



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

CAMPUS D'ALCOI

Análisis del consumo eléctrico por comunidades autónomas a través de modelos econométricos

**MEMORIA PRESENTADA POR:
IGNACIO BLANES ORTIZ**

GRADO DE ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

Convocatoria de defensa: JULIO 2017

Tabla de contenido

1. INTRODUCCIÓN Y MOTIVACIÓN	2
2. OBJETIVOS	5
3. METODOLOGÍA Y TÉCNICAS	7
3.1. METODOLOGÍA	8
3.2. ANÁLISIS UNIVARIANTE	9
3.3. ANÁLISIS BIVARIANTE	11
3.4. MULTICOLINEALIDAD	11
3.5. ANÁLISIS MULTIVARIANTE	12
4. CÁLCULOS	16
4.1. ANÁLISIS UNIVARIANTE	17
4.3. MULTICOLINEALIDAD	40
4.4. ANÁLISIS MULTIVARIANTE	49
4.4.1 PROPOSICION MODELO TEÓRICO Y EXPLICACIÓN DE LOS PARÁMETROS	50
4.4.2. PROPOSICIÓN MODELO AJUSTADO	51
4.4.3. SIGNIFICATIVIDAD	51
4.4.4 NORMALIDAD DE RESIDUOS	53
4.4.5. HETEROCEDASTICIDAD	55
4.4.6 AUTOCORRELACIÓN	56
4.4.7. PUNTOS ANÓMALOS E INFLUYENTES	58
4.5. SOLUCIONES	59
5. PREDICCIONES	63
6. CONCLUSIONES	68
7. LÍNEAS FUTURAS DE INVESTIGACIÓN	70
8. BIBLIOGRAFÍA	72

1. INTRODUCCIÓN Y MOTIVACIÓN

La energía eléctrica, es un tipo de energía secundaria la cual se obtiene a partir de la transformación de energía primaria.

La obtención de energía primaria se puede realizar mediante diferentes tipos de energías como por ejemplo los combustibles crudos, la energía solar, la energía eólica, la energía geotérmica, entre otras. Las energías primarias son aquellas que se extraen de la propia naturaleza.



Ilustración 1, Imagen Energías Primarias.

Tras la obtención de los recursos necesarios para obtener energía primaria, esta se transforma para convertirla en energía secundaria la cual está destinada al consumo directo.

Entre las principales energías secundarias, destacan gasolinas, electricidad, petróleo.

En el caso de la energía eléctrica, tras obtener los recursos de energía primaria necesarios, estos van a centrales transformadoras para convertirla en energía secundaria, en este caso electricidad, la cual mediante las centrales eléctricas se distribuye hasta llegar a nuestras casas, puestos de trabajo, lugares de ocio...

La pregunta es, ¿Puede llegar el día en que esta transformación no sea posible?

Es más que evidente que el planeta viene sufriendo grandes cambios naturales, debido al alto grado de contaminación, de emisiones de CO₂, sobreexplotación, tala de árboles, etc. Lo cual está provocando sequías, aumentos de temperaturas, deshielo, entre otros problemas naturales; cambios que pueden afectar de una manera muy significativa nuestro día a día. Debido a estos cambios, los cuales muchos científicos pronostican que van a ir a peor, pueden hacer que la obtención de energías primarias como se conoce actualmente para posteriormente transformarlas en secundarias se vea mermado lo cual provocaría que la obtención de energía primaria no se podría llevar a cabo (como se ha llevado hasta el momento) por lo que no se podría transformar en energía secundaria. Esta situación está provocando e incluso obligando a la sociedad a buscar alternativas, las cuales encarecen el proceso de obtención de energía primaria lo cual hace que el precio de consumo aumente significativamente.

En la actualidad hay una gran preocupación respecto al consumo eléctrico, y no es para menos, al ser un tipo de energía que depende de la transformación de energía primaria (energía eólica, energía mareomotriz, entre otras) debido a los cambios anteriormente comentados, la falta de lluvias, viento, mareas, está provocando que la obtención de esta sea muy problemático, lo cual hace que se tengan que buscar alternativas, como por ejemplo petróleo, alternativas que son más caras las cuales hacen que el proceso sea más costoso en cuanto a la búsqueda de recursos para transformar se refiere. De ahí viene el problema de las subidas de precios prácticamente.

Por lo que esta es la motivación principal a la hora de realizar, el pensar o la preocupación de que puede llegar el día en el que le demos al interruptor y no funcionar la luz o verse restringido el consumo debido a la falta de recursos.

Por lo que es imprescindible que la sociedad empiece a concienciarse de que llegara un punto en el que no habrá recursos y será todo por un mal o excesivo uso de los recursos.

2. OBJETIVOS

- El principal objetivo a la hora de realizar este proyecto es, si existe algún tipo de relación entre la variable explicada, consumo energético medio, y las variables explicativas, superficie, población, empresas activas, locales activos, PIB, importaciones, exportaciones, horas de sol, renta por hogar, emisiones de CO₂, temperatura media.
- Siguiendo esta línea, la realización de este proyecto va a permitir comprobar la situación de las distintas comunidades autónomas españolas en cuanto al consumo energético, es decir, si es el adecuado, excesivo o insuficiente.

3. METODOLOGÍA Y TÉCNICAS

3.1. METODOLOGÍA

La realización de este proyecto va a consistir en a partir de buscar unas variables, realizar un modelo para posteriormente realizar unas predicciones y finalmente realizar las conclusiones.

Para ello el primer paso va a ser la obtención de la base de datos para posteriormente comenzar con los cálculos. Estos datos se han obtenido de la siguiente manera:

VARIABLE	FUENTE
Superficie	INE
Población	INE
Empresas Activas	INE
Locales Activos	INE
PIB	DATOSMACRO
Exportaciones	IVACE
Importaciones	DATOSMACRO
Temperatura Media	INE
Horas de Sol	INE
Emisiones de CO ₂	MAPAMA
Consumo Energético Medio por CCAA	CERTICALIA
Renta Media por Hogar	INE

Tabla 1, Fuentes de información de las variables.

Una vez obtenidos los datos, se va a proceder a realizar los diferentes cálculos y análisis para posteriormente realizar unas predicciones y obtener finalmente unas conclusiones.

El primero de los cálculos o análisis es el análisis univariante, donde en esta parte se realizará el análisis de cada variable para conocer los máximos y mínimos, comprobar la dispersión, la simetría de los datos, obtener los puntos anómalos y explicar cada variable.

A continuación, se realizará el análisis bivariante donde se establecerá el coeficiente de correlación de cada variable, estudiar la relación de las variables explicativas con la explicada y el grado de esta relación.

Seguidamente se hará el análisis de multicolinealidad.

Tras realizar el análisis de multicolinealidad se procederá hacer el último de los tres análisis, el análisis multivariante donde se validará y estimará el modelo.

Una vez realizados estos análisis se procederá a realizar las predicciones. Estas van a constar de tres tipos de predicciones en las que:

- En la primera se buscará comprobar cómo se ajusta el modelo realizado a la realidad, es decir, se comparará el dato real obtenido en la base de datos con el dato que según el modelo debería ser. Esta predicción se realizará únicamente de una sola comunidad.
- La segunda de las predicciones va a consistir en realizar 3 hipótesis sobre el modelo, en las que variará algún dato de la base de datos (aumentar o disminuir el dato) y se procederá como en la primera predicción, se comparará el dato real de una comunidad autónoma de consumo energético medio con el dato que debería ser según el modelo y así poder comprobar nuevamente si se ajusta bien o no el modelo a la realidad. Estas modificaciones o hipótesis van a ser:

Aumentar la temperatura media.
Aumentar las horas de sol.
Disminuir emisiones de CO₂.

- Por último, en la tercera predicción se va a realizar lo mismo que en la primera predicción, pero en este caso se va a realizar de todas las comunidades autónomas. Se comparará el dato real obtenido en la base de datos con el dato que según el modelo debería de ser para cada comunidad autónoma. En esta predicción se comentará la situación de cada comunidad autónoma, es decir, si tienen un consumo eléctrico adecuado, excesivo o insuficiente.

Para los tres tipos de predicciones se va a calcular el intervalo de confianza para la media y observar si el dato real obtenido en la base de datos se sitúa dentro de ese intervalo, en ese caso el modelo se ajustará bien a la realidad, en caso contrario esa comunidad tendrá un consumo eléctrico excesivo o insuficiente.

Por último, se realizarán las conclusiones donde se plasmarán los resultados obtenidos en la realización de este proyecto.

Para la realización de este proyecto, se había pensado realizarlo de todas las comunidades autónomas que forman el territorio español, debido a la dificultad a la hora de obtener información de las comunidades autónomas de Ceuta y Melilla, se decide eliminarlas de la base de datos.

Antes de meterse de lleno en la realización de los cálculos, se va a proceder a explicar los conceptos clave de cada análisis para la correcta realización de los distintos cálculos que se van a realizar a lo largo del proyecto.

3.2. ANÁLISIS UNIVARIANTE

El análisis univariante se centra en analizar una a una cada variable de manera independiente.

- **Promedio**

Suma de los datos totales, dividido por el número de datos.

$$\bar{x} = \frac{\sum x_i}{n}$$

- **Mediana**

Representa el valor situado en la posición central cuando todos los datos están ordenados de menor a mayor.

Esta puede ser par o impar

Par: El valor de la mediana se obtiene sumando los dos valores situados en la posición central de los datos y se divide entre dos.

Impar: El valor de la mediana es el dato que este situado en el centro del total de los datos.

- **Varianza**

Medida de dispersión que se obtiene de la suma de diferencia entre los valores y el valor medio al cuadrado dividido entre la población. Se expresa como σ^2 .

$$\frac{\sum_{i=1}^n (X_i - \bar{x})^2}{N}$$

- **Desviación Estándar**

Se obtiene realizando la raíz cuadrada de la varianza.

$$\sqrt{\sigma^2}$$

- **Mínimo**

Dato menor de la variable.

- **Máximo**

Dato mayor de la variable.

- **Rango**

Se obtiene a partir de la diferencia entre el máximo y mínimo.

$$\text{Máximo} - \text{Mínimo}$$

- **Cuartil**

Valores que dividen los datos ordenados de menor a mayor en cuatro partes, son cuatro:

Primer Cuartil (Q1): Representa el 25% de los datos.

Segundo Cuartil (Q2): Representa el 50% de los datos, además coincide con la mediana.

Tercer Cuartil (Q3): Representa el 75% de los datos.

Cuarto Cuartil (Q4): Representa el 100% de los datos. Este coincide con el máximo.

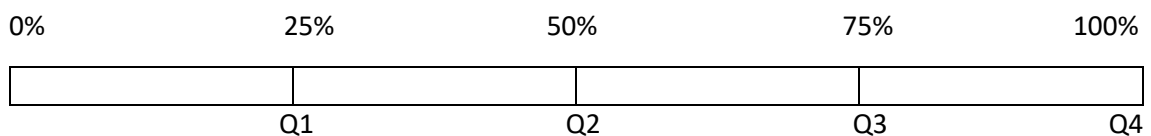


Ilustración 2, Cuartiles.

- **Sesgo**

Es una unidad de medida que muestra la simetría o asimetría de los datos.

- **Curtosis**

Mide el grado de concentración de los valores de la muestra.

- **Gráfico de caja y bigotes**

Muestra visualmente la distribución de los datos, el cual consta de una caja principal en la que aparece la mediana y media. A los lados están los bigotes, al final de los bigotes se sitúan el máximo y mínimo. En caso de existir datos no homogéneos se situarían fuera de los bigotes, a los cuales denominaríamos como punto anómalo.

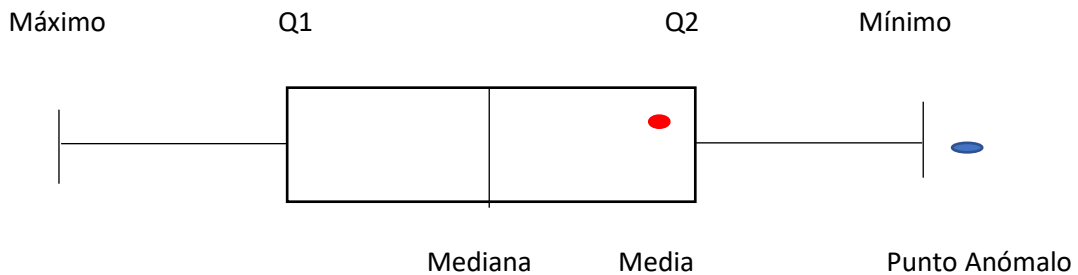


Ilustración 3, Gráfico de caja y bigotes.

- **Punto anómalo**

Dato no homogéneo el cual no se comporta como el resto de los datos.

3.3. ANÁLISIS BIVARIANTE

El análisis bivalente mide la relación entre dos variables.

- **Gráficos XY**

Representación gráfica de los datos en el que se representa la variable explicada frente la variable explicativa.

- **Coefficiente de Correlación**

Coefficiente que muestra la relación lineal que existe entre dos variables. Se conoce como r . Esta puede ser:

Si $r > 0,7 \rightarrow$ Relación muy elevada.

Si $0,5 < r \leq 0,7 \rightarrow$ Relación moderada.

Si $0,3 < r \leq 0,5 \rightarrow$ Relación baja.

Si $0 < r \leq 0,3 \rightarrow$ Relación muy baja.

Será positiva cuando la pendiente de la recta sea ascendente.

Será negativa cuando la pendiente de la recta sea descendente.

3.4. MULTICOLINEALIDAD

Este problema se da cuando dos o más variables explicativas están muy relacionadas entre ellas y sus valores son semejantes. Para comprobar este problema se van a realizar tres pruebas:

Matriz de correlación

En esta prueba se observará los coeficientes por debajo de la diagonal principal de la matriz, si algún coeficiente en valor absoluto es superior a 0,7 será cuando haya problemas de multicolinealidad.

Matriz inversa de correlación

En este caso se observará únicamente la diagonal principal de la matriz y si algún coeficiente es superior a 10, la variable correspondiente a ese coeficiente generara problemas de multicolinealidad.

Índice de acondicionamiento (I.C.)

Por último, tras calcular el índice de acondicionamiento el cual se calcula de la siguiente forma, se deberá tener en cuenta lo siguiente:

$$I.C. = \sqrt{\frac{Autovalor_{MAX}}{Autovalor_{MIN}}}$$

Si $0 \leq I.C. < 10 \rightarrow$ No existe problema de multicolinealidad.

Si $10 \leq I.C. < 20 \rightarrow$ Existe multicolinealidad moderada.

Si $20 \leq I.C. \rightarrow$ Existe multicolinealidad moderada.

Para que no haya problemas de multicolinealidad, tienen que dar negativo las tres pruebas, es decir, que en ninguna prueba se observen problemas de multicolinealidad. En caso de haber problemas de multicolinealidad, este se tiene que resolver. En caso de existir, habrá dos criterios a la hora de resolver el problema, estos son: (Estos criterios se usarán utilizando la matriz inversa de correlaciones)

- Criterio 1: En caso de que en la matriz inversa de correlaciones hubiesen más de un coeficiente superior a 10, lo primero sería observar si la diferencia entre los dos coeficientes más altos es superior a 1, en este caso se eliminara de la base de datos la variable correspondiente al coeficiente más alto, siempre y cuando la diferencia entre el coeficiente mayor y el segundo más alto sea mayor de 1.
- Criterio 2: En caso de que la diferencia entre los coeficientes más altos sea menor a 1, se procederá a realizar dos regresiones diferentes utilizando en cada una de ellas una variable cada vez de las que generan el problema. Tras realizar las regresiones se eliminará la variable que menor R^2 genere.

3.5. ANÁLISIS MULTIVARIANTE

El análisis multivariante es donde se lleva a cabo la estimación y validación del modelo.

- Modelo Teórico

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_n * X_n + U$$

- **Significatividad**

Contraste de hipótesis mediante el cual se comprueba la significatividad del modelo como de cada parámetro.

Para el modelo:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_1: \text{Al menos uno es } \neq 0$$

Para cada parámetro:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

- **Análisis normalidad de residuos**

Conjunto de pruebas que se llevan a cabo para comprobar la normalidad de los residuos, estas son:

Histograma

Representación gráfica en la que se observa la distribución que sigue la muestra. Si la distribución tiene forma de campana de Gauss los residuos son normales, en caso de no tener esta forma, los residuos no son normales.

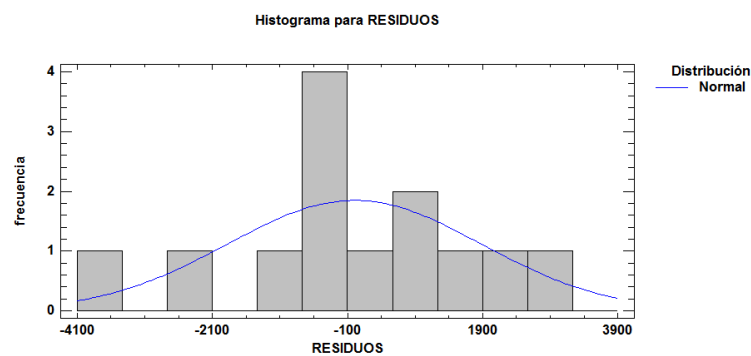


Ilustración 4, Histograma.

Gráfico probabilístico

Grafico en el que se observa la normalidad de los residuos, para ello tienen que estar cerca de la línea de normalidad, en caso de estar más dispersos de esta línea no serán normales los residuos.

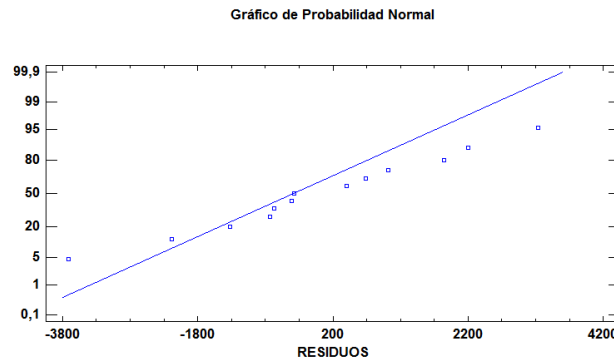


Ilustración 5, Gráfico Probabilístico.

Test de normalidad de residuos

Contraste que concluye la normalidad o no de los residuos. El cuál es el siguiente:

$$\left. \begin{array}{l} H_0: \text{Los residuos son normales} \\ H_1: \text{Los residuos no son normales} \end{array} \right\} \alpha = 0,05$$

- Análisis heterocedasticidad

El modelo tendrá problemas de heterocedasticidad cuando la varianza de las perturbaciones no es constante. Este se detecta a través de otro contraste de hipótesis:

$$\left. \begin{array}{l} H_0: \sigma^2: \text{Constante} \\ H_1: \sigma^2: \text{No es constante} \end{array} \right\} \alpha = 0,05$$

- Autocorrelación

Habrà problema de autocorrelación en el modelo cuando los residuos estén correlacionados. Para detectar este problema se realizará lo siguiente:

Test Durbin-Watson

Contraste que permite detectar si el modelo tiene problema de autocorrelación.

$$\left. \begin{array}{l} H_0: \rho_1 = 0 \text{ No hay autocorrelación.} \\ H_1: \rho_1 \neq 0 \text{ Hay autocorrelación.} \end{array} \right\} \alpha = 0,05$$

FAS y FAP

Gráficos en los que se observa si hay problema de autocorrelación. Para que no haya este problema, en este gráfico ningún factor de autocorrelación debe superar los límites establecidos.

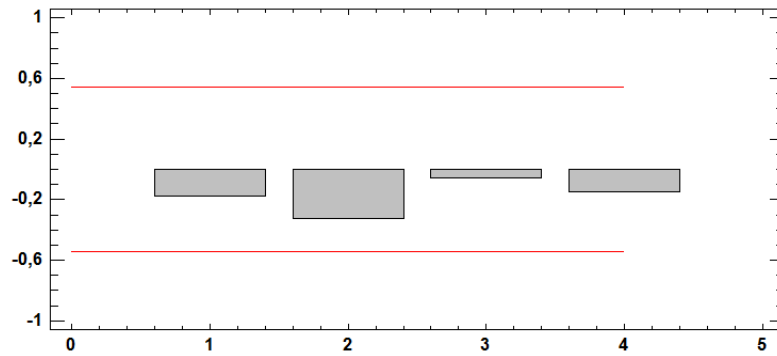


Ilustración 6, FAS y FAP.

- Puntos influyentes

Valores que ofrece el modelo los cuales son importantes en lo que al ajuste del modelo se refiere.

4. CÁLCULOS

4.1. ANÁLISIS UNIVARIANTE

El primero de los análisis a realizar va a ser el Análisis Univariante. El objetivo de este análisis es conocer los máximos y mínimos de cada variable de la base de datos, observar la dispersión de los datos, es decir, si están poco o muy dispersos. Para ello se comparará la desviación estándar con el promedio y así poder obtener conclusiones de cada variable. Para ello se deberá tener en cuenta que:

Si la Desviación Estándar \leq Promedio \rightarrow Los datos están poco dispersos.

Si la Desviación Estándar $>$ Promedio \rightarrow Los datos están muy dispersos.

Al igual que servirá para determinar la simetría o asimetría de los datos, esto se llevará a cabo observando el gráfico de caja y bigotes de cada variable.

Por último, este análisis también va a ayudar para determinar los puntos anómalos de cada variable.

Al ser este el primero de los análisis a realizar se va a aprovechar para comentar que mide cada variable y cuál es su unidad de medida.

También indicar que las siglas de cada variable hacen referencia a la abreviatura de la misma variable, las cuales serán utilizadas para el análisis de la multicolinealidad.

- **SUPERFICIE (SUP)**

Esta variable mide la extensión de cada comunidad autónoma de España medida en km².

Resumen de Estadísticos

PROMEDIO	29481,5
MEDIANA	10604,0
DESVIACIÓN ESTÁNDAR	30629,7
MÍNIMO	4992,0
MÁXIMO	94222,0
RANGO	89230,0
CUARTIL INFERIOR	7234,0
CUARTIL SUPERIOR	41634,0
SESGO	1,22362
CURTOSIS	0,209742

Tabla 2, Resumen de Estadísticos Variable Superficie

Tras haber calculado el resumen de estadísticos se obtiene que:

Máximo: 94222 correspondiente a la comunidad autónoma de Castilla y León.

Mínimo: 4992,0 correspondiente a la comunidad autónoma de Islas Baleares.

Puntos anómalos

Gráfico de Caja y Bigotes

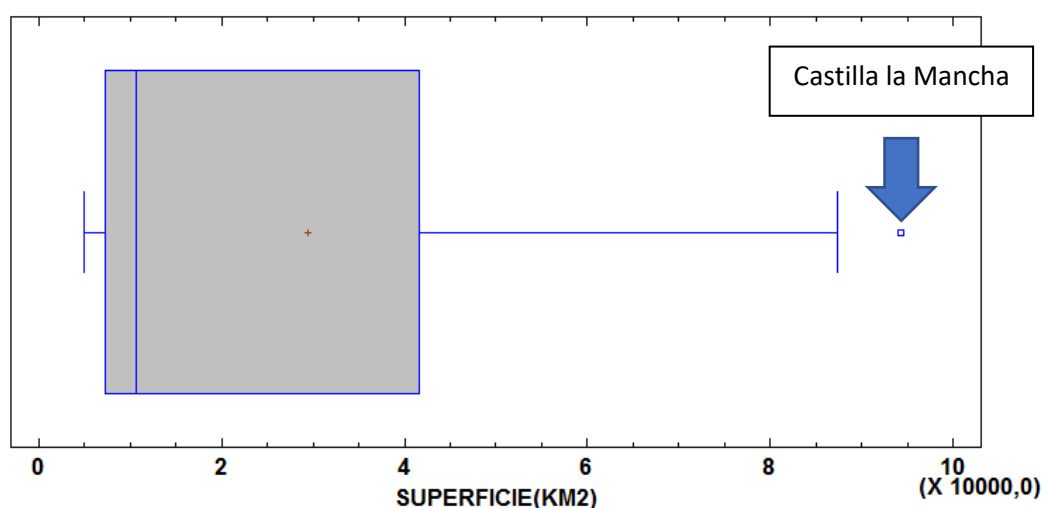


Ilustración 7, Gráfico de Caja y Bigotes Variable Superficie

En este caso los datos están muy dispersos y siguen una distribución asimétrica a la derecha.

- **POBLACIÓN (POB)**

Esta variable mide la población total de cada comunidad autónoma medida en millones de personas.

Resumen estadístico

PROMEDIO	2,48742
MEDIANA	2,079
DESVIACIÓN ESTÁNDAR	2,38094
MÍNIMO	0,319
MÁXIMO	8,402
RANGO	8,083
CUARTIL INFERIOR	1,062
CUARTIL SUPERIOR	2,749
SESGO	1,62009
CURTOSIS	1,96195

Tabla 3, Resumen de Estadísticos Variable Población

Tras haber calculado el resumen de estadísticos se obtiene que:

Máximo: 8,402 correspondiente a la comunidad autónoma de Andalucía.

Mínimo: 0,319 correspondiente a la comunidad autónoma de La Rioja.

Puntos anómalos

Gráfico de Caja y Bigotes

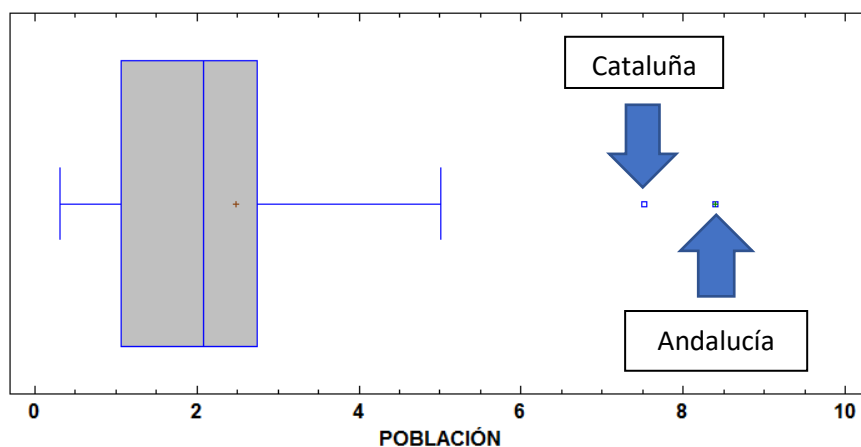


Ilustración 8, Gráfico de Caja y Bigotes Variable Población.

En este caso los datos están poco dispersos y siguen una distribución asimétrica a la derecha.

- EMPRESAS ACTIVAS (EA)

Esta variable mide el número de empresas totales activas por cada comunidad autónoma, medida en miles de empresas.

Resumen estadístico

PROMEDIO	187002
MEDIANA	125786
DESVIACIÓN ESTÁNDAR	178096
MÍNIMO	23083,0
MÁXIMO	584369
RANGO	561286
CUARTIL INFERIOR	67451,0
CUARTIL SUPERIOR	195595
SESGO	1,3438
CURTOSIS	0,501157

Tabla 4, Resumen de Estadísticos Variable Empresas Activas.

Tras haber calculado el resumen de estadísticos se obtiene que:

Máximo: 584369 correspondiente a la comunidad autónoma de Cataluña.

Mínimo: 23083 correspondiente a la comunidad autónoma de La Rioja.

Puntos anómalos

Gráfico de Caja y Bigotes

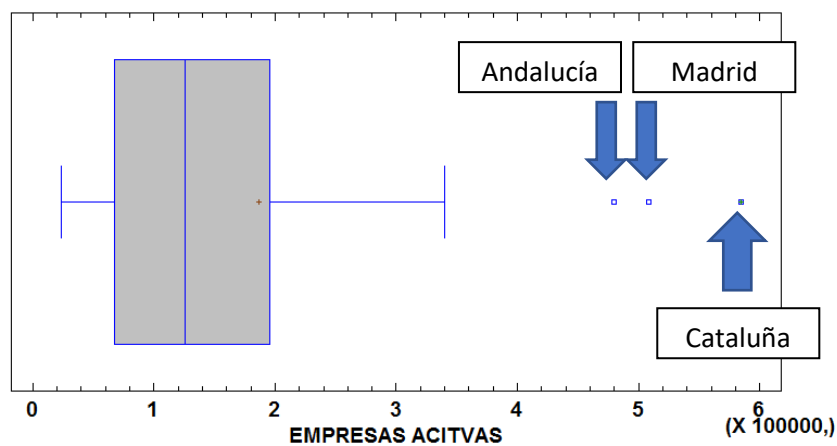


Ilustración 9, Gráfico de Caja y Bigotes Variable Empresas Activas.

En este caso los datos están poco dispersos y siguen una distribución asimétrica a la derecha.

- LOCALES ACTIVOS (LA)

Esta variable mide el número de locales activos por comunidad autónoma, medida en miles de locales.

Resumen de Estadísticos

PROMEDIO	211166
MEDIANA	144913
DESVIACIÓN ESTÁNDAR	198929
MÍNIMO	26421,0
MÁXIMO	656832
RANGO	630411
CUARTIL INFERIOR	76754,0
CUARTIL SUPERIOR	220314
SESGO	1,33159
CURTOSIS	0,482514

Tabla 5, Resumen de Estadísticos Variable Locales Activos.

Tras haber calculado el resumen de estadísticos se obtiene que:

Máximo: 656832 correspondiente a la comunidad autónoma de Cataluña.

Mínimo: 26421 correspondiente a la comunidad autónoma de La Rioja.

Puntos anómalos

Gráfico de Caja y Bigotes

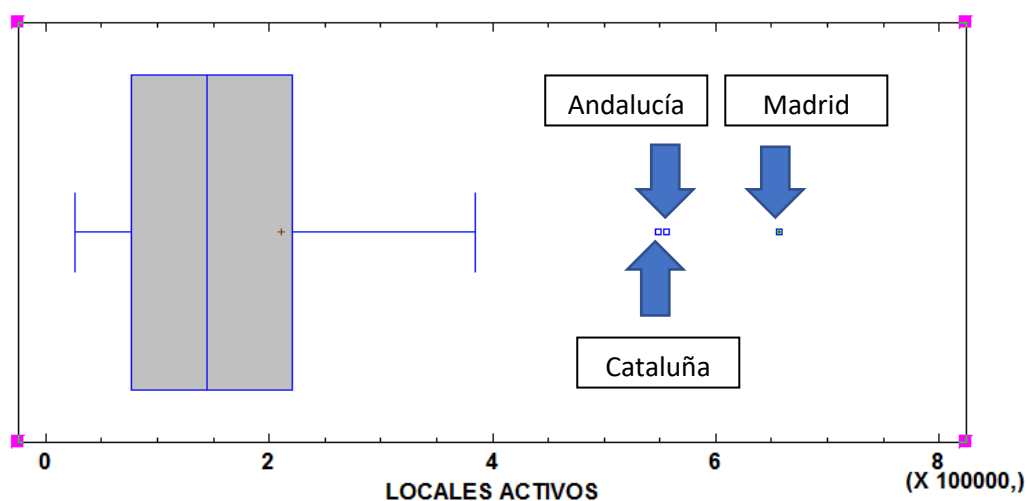


Ilustración 10, Gráfico de Caja y Bigotes Variable Locales Activos.

En este caso los datos están poco dispersos y siguen una distribución simétrica pese a observarse 3 puntos anómalos a la derecha.

- **PIB (PIB)**

Mide la actividad productiva de un país, medida en miles de euros.

Resumen estadístico

PROMEDIO	61020,4
MEDIANA	36463,0
DESVIACIÓN ESTÁNDAR	60819,3
MÍNIMO	7727,0
MÁXIMO	197004,
RANGO	189277,
CUARTIL INFERIOR	20793,0
CUARTIL SUPERIOR	63394,0
SESGO	1,53442
CURTOSIS	1,28047

Tabla 6, Resumen de Estadísticos Variable PIB.

Tras haber calculado el resumen de estadísticos se obtiene que:

Máximo: 197004 correspondiente a la comunidad autónoma de Cataluña.

Mínimo: 7727 correspondiente a la comunidad autónoma de La Rioja.

Puntos anómalos

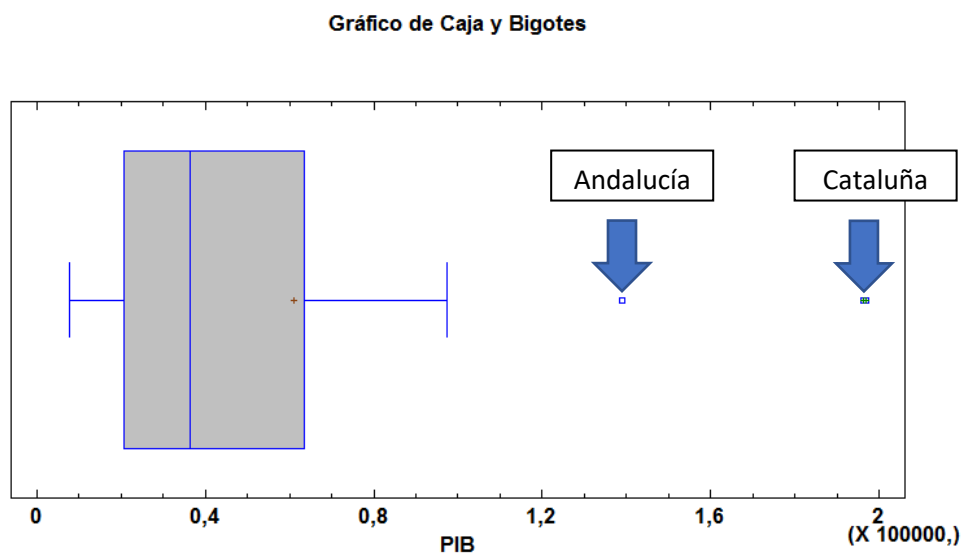


Ilustración 11, Gráfico de Caja y Bigotes Variable PIB.

En este caso los datos están poco dispersos y siguen una distribución asimétrica a la derecha a la derecha.

- EXPORTACIONES (EXP)

Esta variable mide el número de exportaciones que realiza cada comunidad autónoma, medida en miles de exportaciones.

Resumen estadístico

PROMEDIO	14065,8
MEDIANA	9382,4
DESVIACIÓN ESTÁNDAR	15181,8
MÍNIMO	924,0
MÁXIMO	60291,2
RANGO	59367,2
CUARTIL INFERIOR	2546,5
CUARTIL SUPERIOR	22501,0
SESGO	1,9026
CURTOSIS	4,53257

Tabla 7, Resumen de Estadísticos Variable Exportaciones.

Tras haber calculado el resumen de estadísticos se obtiene que:

Máximo: 60291.2 correspondiente a la comunidad autónoma de Cataluña.

Mínimo: 924 correspondiente a la comunidad autónoma de Islas Baleares.

Puntos anómalos

Gráfico de Caja y Bigotes

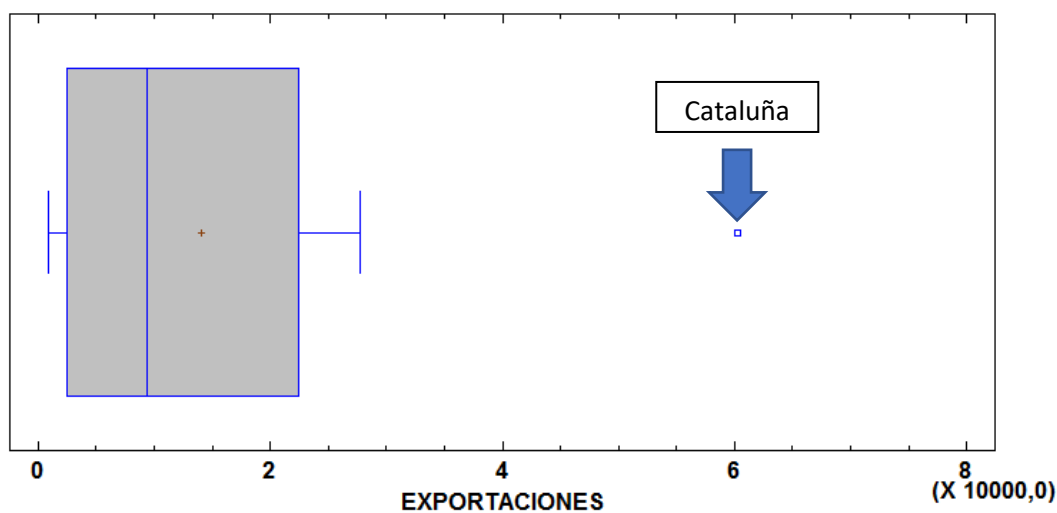


Ilustración 12, Gráfico de Caja y Bigotes Variable Exportaciones.

En este caso los datos están muy dispersos y siguen una distribución simétrica y se observa la presencia de Cataluña como punto anómalo.

- IMPORTACIONES (IMP)

Esta variable mide el número de importaciones realizadas por cada comunidad autónoma, medida en miles de importaciones.

Resumen estadístico

PROMEDIO	15423,7
MEDIANA	8552,9
DESVIACIÓN ESTÁNDAR	19567,2
MÍNIMO	980,4
MÁXIMO	72220,5
RANGO	71240,1
CUARTIL INFERIOR	3342,0
CUARTIL SUPERIOR	17131,4
SESGO	2,02759
CURTOSIS	3,96425

Tabla 8, Resumen de Estadísticos Variable Importaciones.

Tras haber calculado el resumen de estadísticos se obtiene que:

Máximo: 72220,5 correspondiente a la comunidad autónoma de Cataluña.

Mínimo: 980,4 correspondiente a la comunidad autónoma de Extremadura.

Puntos anómalos

Gráfico de Caja y Bigotes

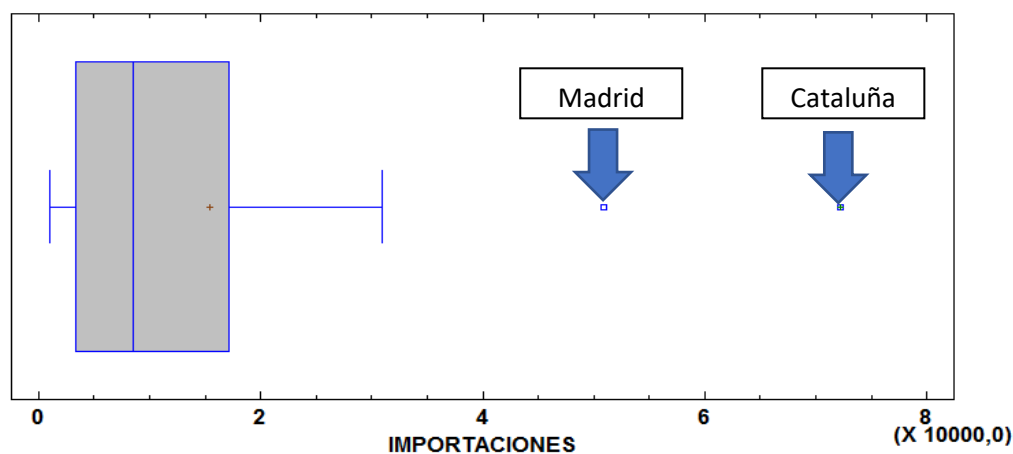


Ilustración 13, Gráfico de Caja y Bigotes Variable Importaciones.

En este caso los datos están muy dispersos y siguen una distribución simétrica con 2 comunidades con un comportamiento distinto al resto.

- **TEMPERATURA MEDIA (TEMP)**

Esta variable mide la temperatura media de cada comunidad autónoma, medida en grados centígrados.

Resumen estadístico

PROMEDIO	16,2457
MEDIANA	15,85
DESVIACIÓN ESTÁNDAR	2,11698
MÍNIMO	12,48
MÁXIMO	20,1
RANGO	7,62
CUARTIL INFERIOR	14,6
CUARTIL SUPERIOR	17,967
SESGO	0,190083
CURTOSIS	-0,793425

Tabla 9, Resumen de Estadísticos Variable Temperatura Media.

Tras haber calculado el resumen de estadísticos se obtiene que:

Máximo: 20,1 correspondiente a la comunidad autónoma de Murcia.

Mínimo: 12,48 correspondiente a la comunidad autónoma de Castilla y León.

Puntos anómalos

Gráfico de Caja y Bigotes

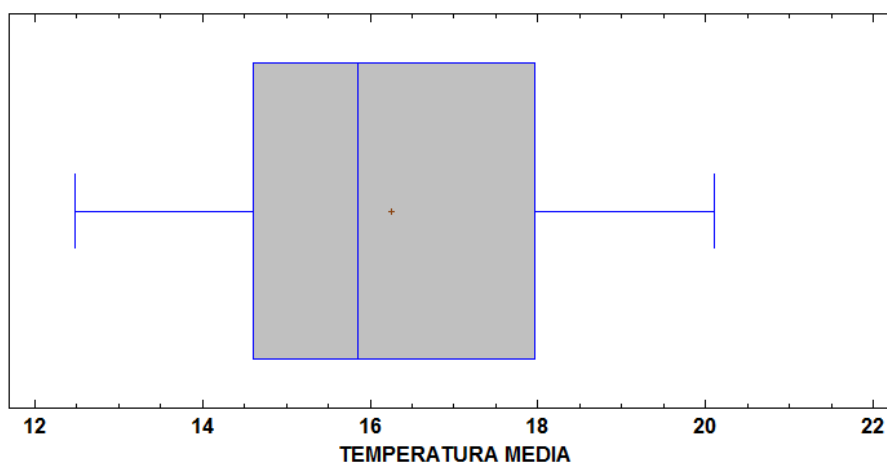


Ilustración 14, Gráfico de Caja y Bigotes Temperatura Media.

En este caso los datos están poco dispersos y siguen una distribución simétrica.

- HORAS DE SOL (HS)

Esta variable mide el número de horas de sol de cada comunidad autónoma, medida en miles de horas.

Resumen estadístico

PROMEDIO	2513,22
MEDIANA	2635,7
DESVIACIÓN ESTÁNDAR	545,798
MÍNIMO	1515,0
MÁXIMO	3418,5
RANGO	1903,5
CUARTIL INFERIOR	2086,0
CUARTIL SUPERIOR	2997,0
SESGO	-0,252861
CURTOSIS	-0,831703

Tabla 10, Resumen de Estadísticos Variable Horas de Sol.

Tras haber calculado el resumen de estadísticos se obtiene que:

Máximo: 3418,5 correspondiente a la comunidad autónoma de Islas Canarias.

Mínimo: 1515 correspondiente a la comunidad autónoma de Cantabria.

Puntos anómalos

Gráfico de Caja y Bigotes

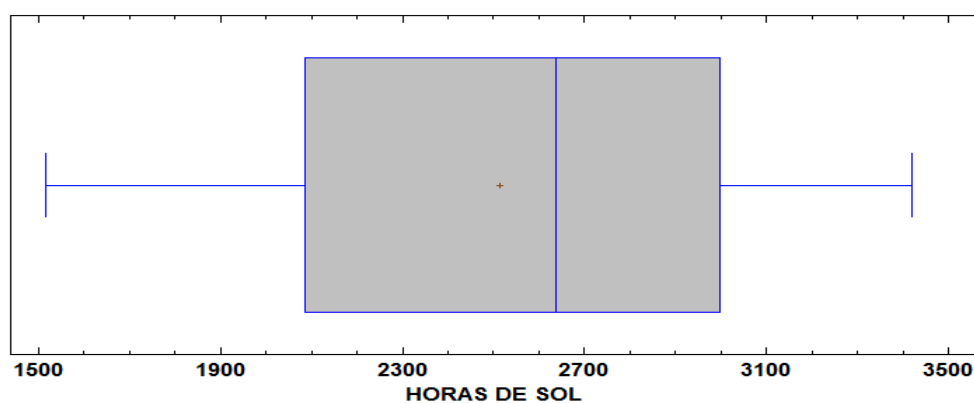


Ilustración 15, Gráfico de Caja y Bigotes Horas de Sol.

En este caso los datos están poco dispersos y siguen una distribución simétrica.

- **EMISIONES CO₂ (CO₂)**

Esta variable mide el número de emisiones de CO₂ de cada comunidad autónoma, medida en KT CO₂/año.

Resumen estadístico

PROMEDIO	19313,3
MEDIANA	16662,0
DESVIACIÓN ESTÁNDAR	13701,3
MÍNIMO	1924,0
MÁXIMO	50623,0
RANGO	48699,0
CUARTIL INFERIOR	8180,0
CUARTIL SUPERIOR	25696,0
SESGO	0,937368
CURTOSIS	0,392392

Tabla 11, Resumen de Estadísticos Variables Emisiones CO₂.

Tras haber calculado el resumen de estadísticos se obtiene que:

Máximo: 50623 correspondiente a la comunidad autónoma de Andalucía.

Mínimo: 1924 correspondiente a la comunidad autónoma de La Rioja.

Puntos anómalos

Gráfico de Caja y Bigotes

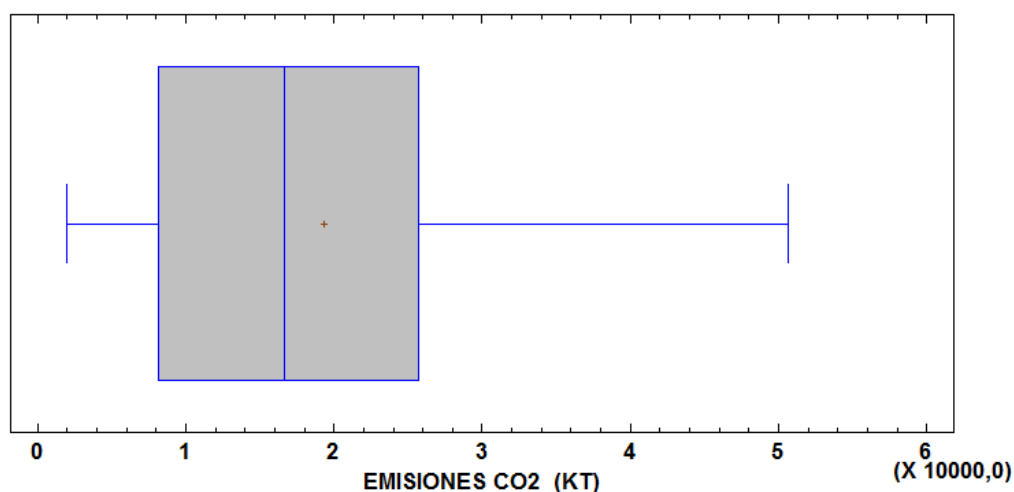


Ilustración 16, Gráfico de Caja y Bigotes Emisiones CO₂.

En este caso los datos están poco dispersos y siguen una distribución asimétrica a la derecha.

- RENTA MEDIA POR HOGAR (RENT)

Esta variable mide la renta media por hogar de cada comunidad autónoma, medida en miles de euros.

PROMEDIO	25898,4
MEDIANA	25623,0
DESVIACIÓN ESTÁNDAR	4476,39
MÍNIMO	19364,0
MÁXIMO	34240,0
RANGO	14876,0
CUARTIL INFERIOR	22271,0
CUARTIL SUPERIOR	29120,0
SESGO	0,4731
CURTOSIS	-0,834626

Tabla 12, Resumen de Estadísticos Variable Renta Media por Hogar.

Tras haber calculado el resumen de estadísticos se obtiene que:

Máximo: 34240 correspondiente a la comunidad autónoma de País Vasco

Mínimo: 19364 correspondiente a la comunidad autónoma de Extremadura

Gráfico de Caja y Bigotes

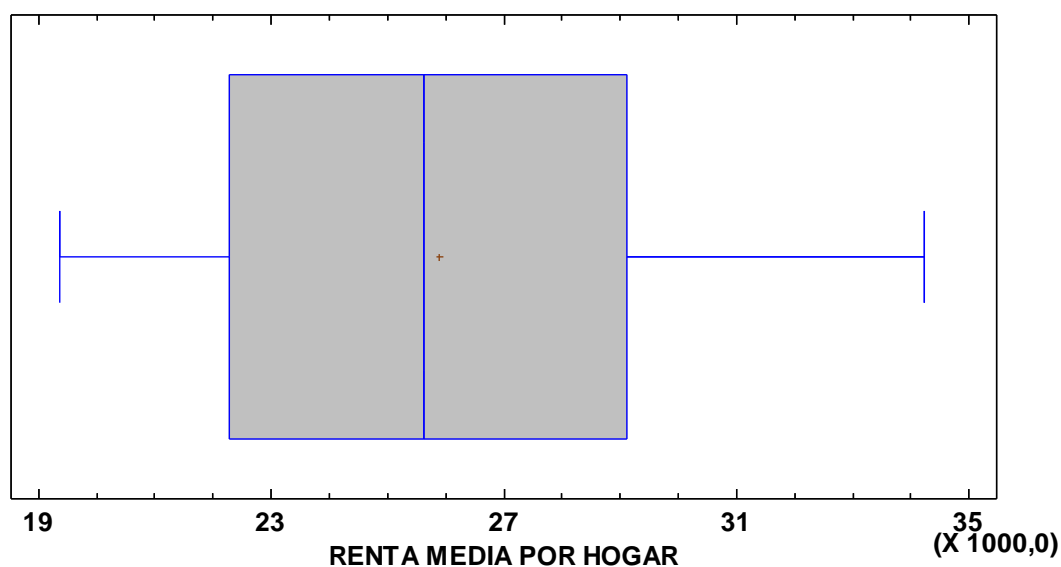


Ilustración 17, Gráfico de Caja y Bigotes Renta por Hogar.

En este caso los datos están muy dispersos y siguen una distribución asimétrica a la derecha.

Tras haber realizado el análisis univariante se observa que las variables Cataluña y Madrid son muy conflictivas ya que no siguen el comportamiento de las otras Comunidades Autónomas saliendo como puntos anómalos en varios análisis, por lo que se decide eliminarlas de la base de datos. Ya que estas comunidades más adelante darán otro tipo de problemas, por lo que lo más idóneo es eliminarlas ya para así evitar problemas futuros.

4.2. ANÁLISIS BIVARIANTE

Llegados a este punto se va a proceder a realizar el análisis bivalente para así determinar el coeficiente de correlación de cada variable.

Al igual que se va a estudiar el tipo de relación que tiene la variable Y con las variables independientes, es decir, si es directa o indirecta como el grado, muy baja, baja, moderada o muy elevada.

- Superficie vs consumo energético medio por CCAA

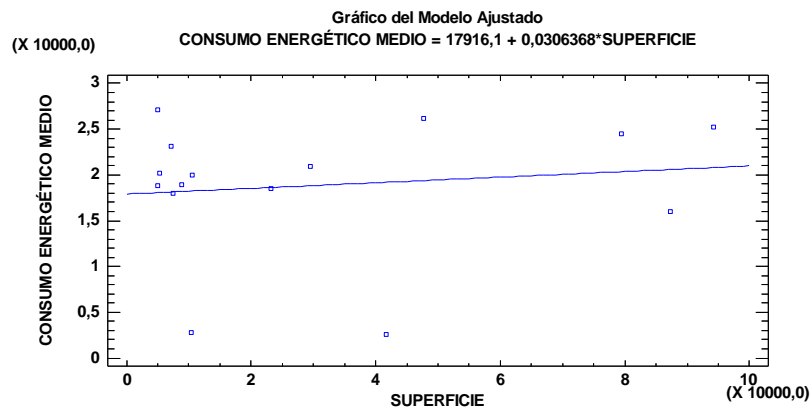


Ilustración 18, Superficie vs consumo energético medio por CCAA.

Una vez realizado el análisis bivalente se obtiene que el coeficiente de correlación es de 0,134128, por lo que la relación es directa y muy baja.

- Población vs consumo energético medio por CCAA

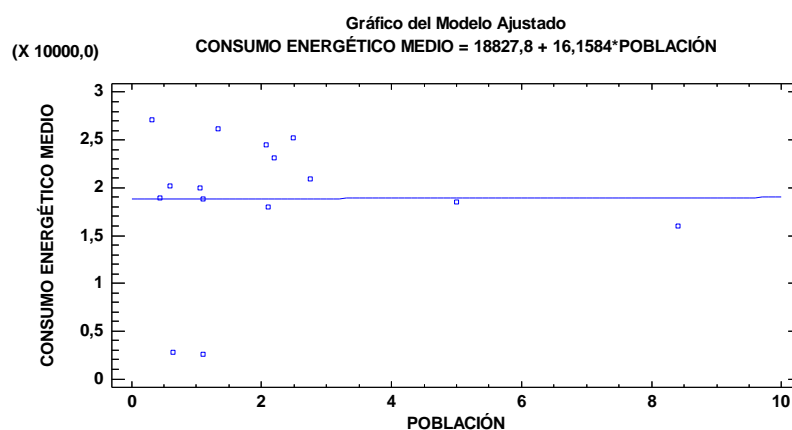


Ilustración 19, Población vs consumo energético medio por CCAA.

Tras realizar el análisis bivalente se obtiene un coeficiente de correlación de 0,00466576, lo cual hace indicar que la relación es muy baja.

- **Empresas activas vs consumo energético medio por CCAA**

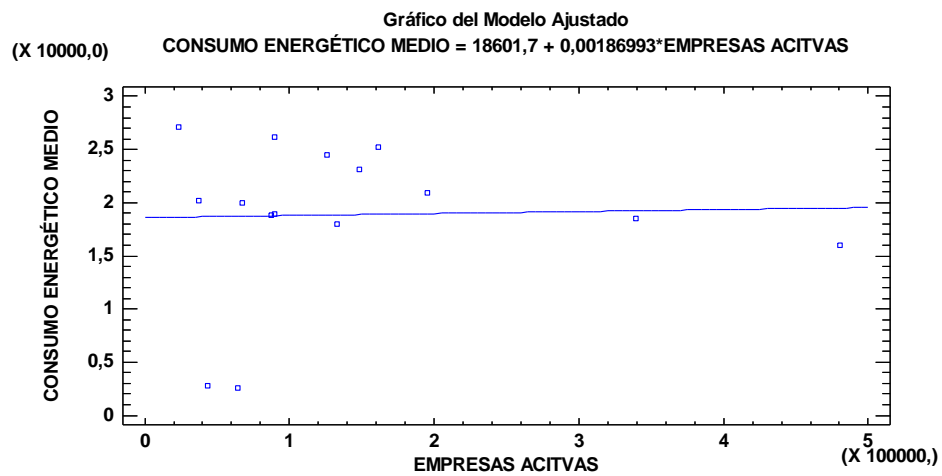


Ilustración 20, Empresas Activas vs consumo energético medio por CCAA.

Una vez realizado el análisis bivalente se obtiene que el coeficiente correlación es de 0,031361, por lo que la relación es baja.

- **Locales activos vs consumo energético medio por CCAA**

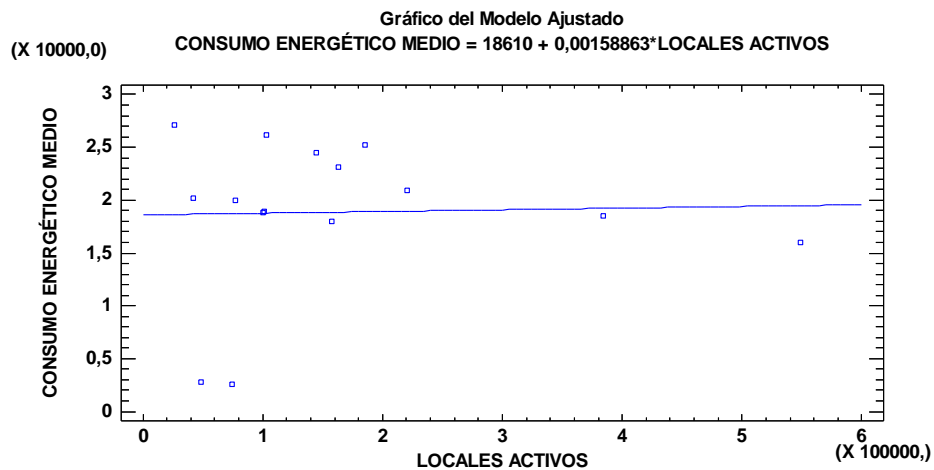


Ilustración 21, Locales Activos vs consumo energético medio por CCAA.

Tras realizar el análisis bivalente se obtiene que el coeficiente correlación es de 0,0303288, por lo que la relación en este caso es baja.

- **PIB vs consumo energético medio por CCAA**

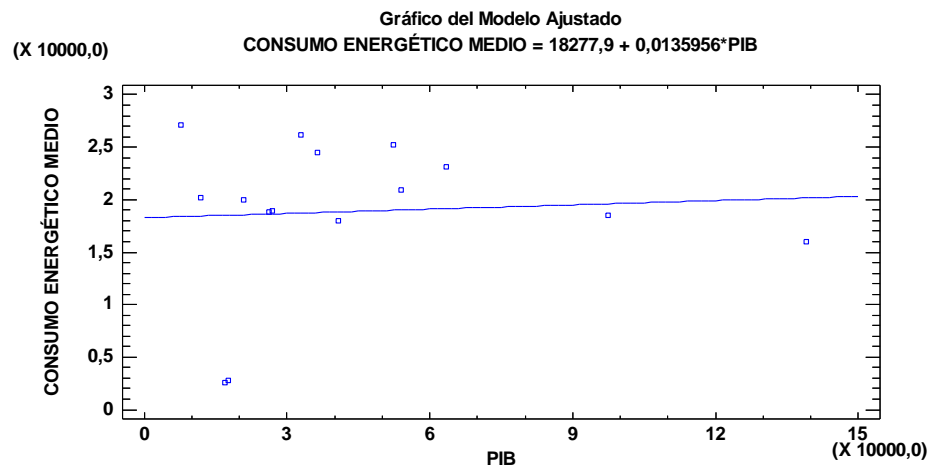


Ilustración 22, PIB vs consumo energético medio por CCAA.

Al realizar el análisis bivariante y obtener un coeficiente correlación de 0,0656359, por lo que se puede observar que la relación es muy baja.

- **Exportaciones vs consumo energético medio por CCAA**

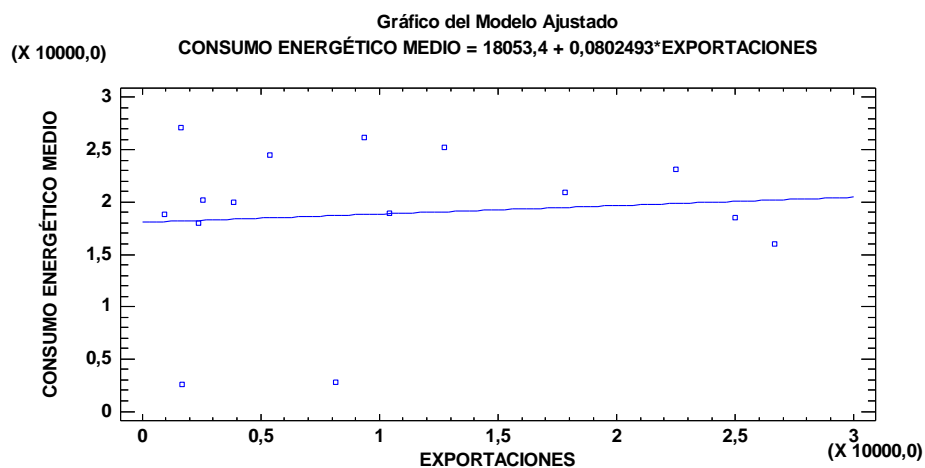


Ilustración 23, Exportaciones vs consumo energético medio por CCAA.

En este caso al realizar el análisis bivariante se obtiene que el coeficiente correlación es de 0,0981177, por lo que la relación es directa y muy baja.

- **Importaciones vs consumo energético medio por CCAA**

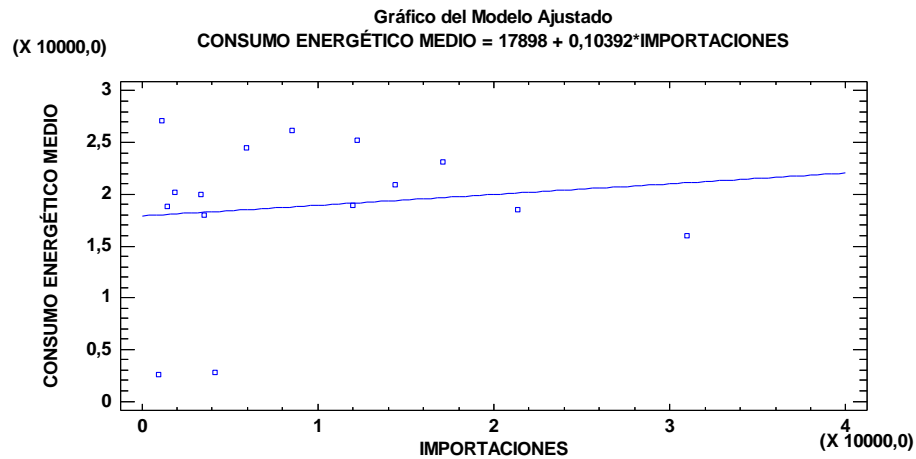


Ilustración 24, Importaciones vs consumo energético medio por CCAA.

Una vez realizado el análisis bivalente se obtiene que el coeficiente de correlación es de 0,124311, por lo que la relación es directa y muy baja.

- **Temperatura media vs consumo energético medio por CCAA**

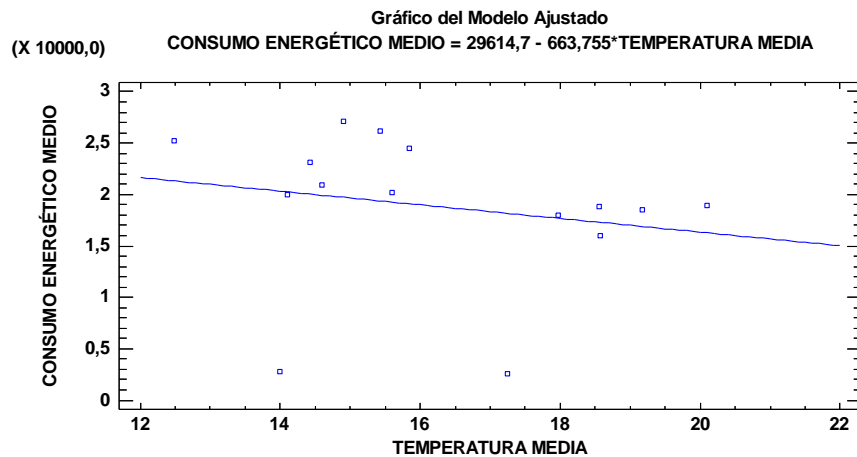


Ilustración 25, Temperatura Media vs consumo energético medio por CCAA.

Tras realizar el análisis bivalente, el programa nos calcula que el coeficiente de correlación es de -0,203816, por lo que la relación es indirecta y baja.

- **Horas de sol vs consumo energético medio por CCAA**

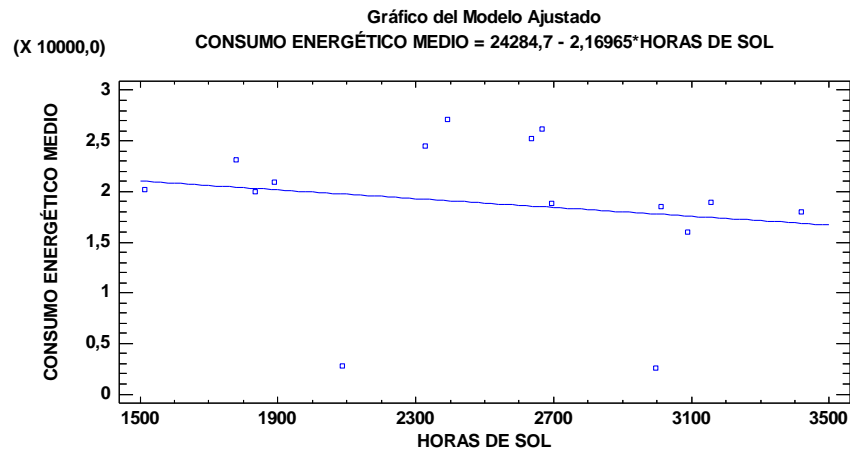


Ilustración 26, Horas de Sol vs consumo energético medio por CCAA.

Una vez realizado el análisis bivariante se obtiene que el coeficiente de correlación es de -0,172018, por lo que la relación es indirecta y muy baja.

- **Emisiones CO₂ vs consumo energético medio por CCAA**

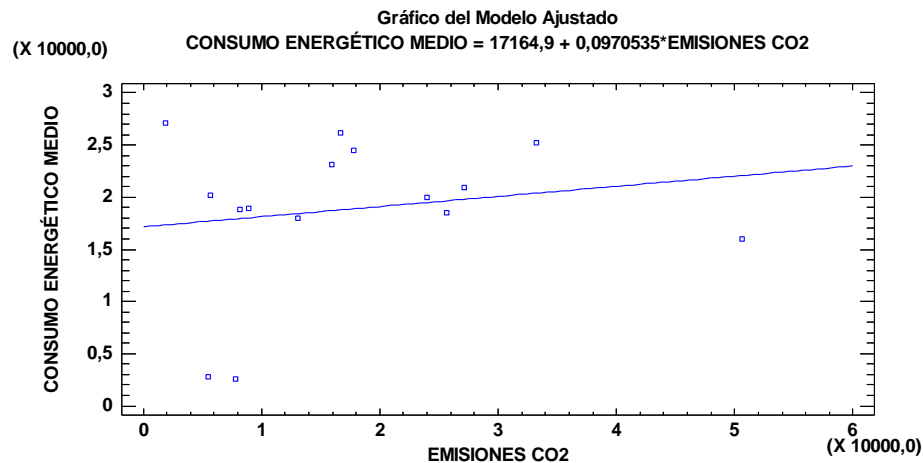


Ilustración 27, Emisiones CO₂ vs consumo energético medio por CCAA.

Al realizar el análisis bivariante se obtiene que el coeficiente de correlación es de 0,17145, por lo que la relación es directa y muy baja.

- Renta por hogar vs consumo energético medio por CCAA

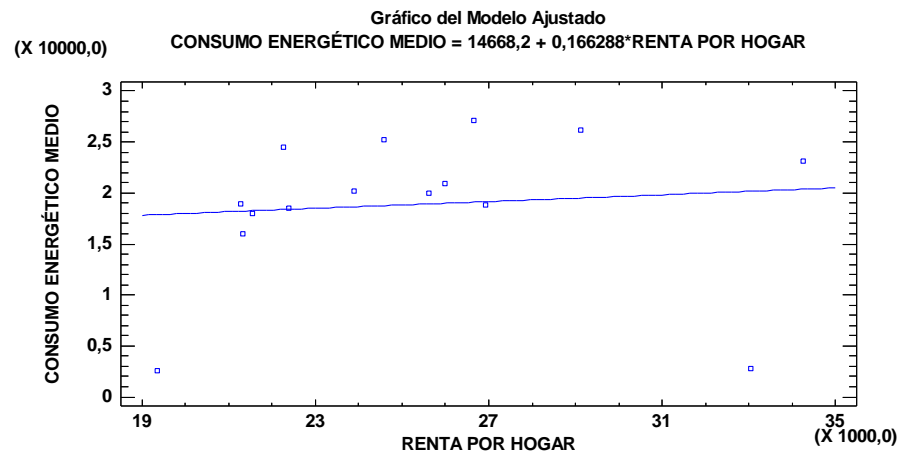


Ilustración 28, Renta por Hogar vs consumo energético medio por CCAA.

Una vez realizado el análisis bivalente se obtiene que el coeficiente de correlación es de 0,0979221, por lo que la relación es directa y muy baja.

Tras realizar el análisis bivalente, se observa que tanto Extremadura como Navarra se comportan de forma anómala al resto de variables, por lo que se decide eliminar estas de la base de datos y volver a realizar el análisis bivalente para comprobar si varían los resultados.

- **Superficie vs consumo energético medio por CCAA**

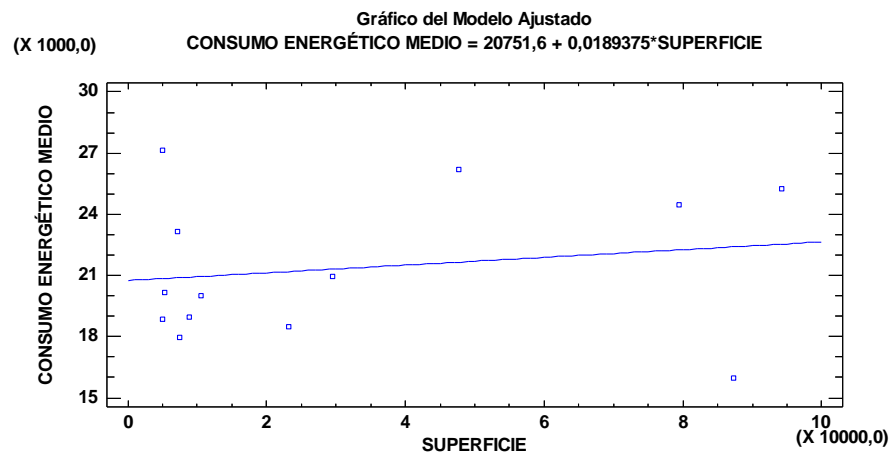


Ilustración 29, Superficie vs consumo energético medio por CCAA.

Una vez realizado el análisis bivalente se obtiene que el coeficiente de correlación es de 0,182897 por lo que la relación es directa y muy baja.

- **Población vs consumo energético medio por CCAA**

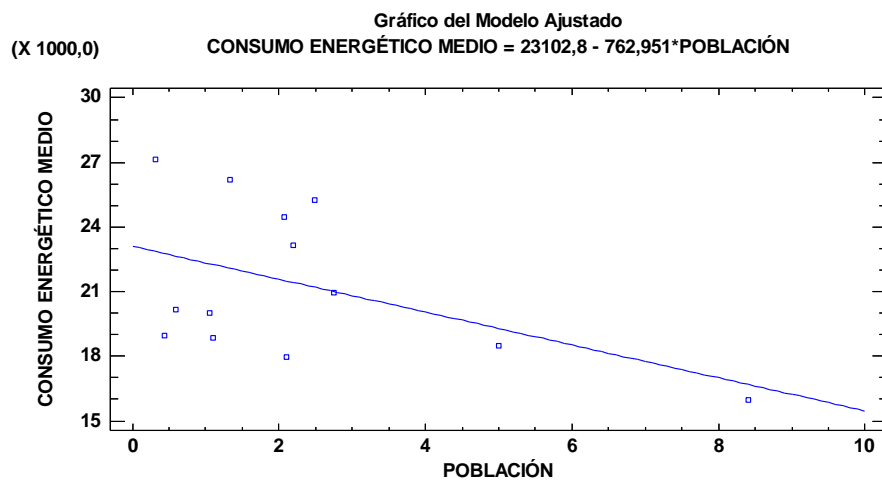


Ilustración 30, Población vs consumo energético medio por CCAA.

Tras realizar el análisis bivalente se obtiene que el coeficiente de correlación es de -0,480843, por lo que la relación es de indirecta y baja.

- **Empresas activas vs consumo energético medio por CCAA**

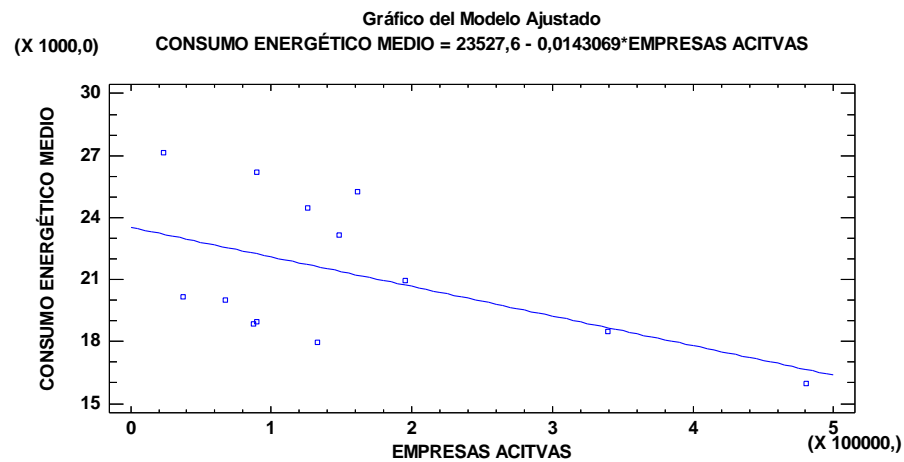


Ilustración 31, Empresas Activas vs consumo energético medio por CCAA.

Una vez realizado el análisis bivalente se obtiene que el coeficiente de correlación es de -0,517439 por lo que la relación es indirecta y moderada.

- **Locales activos vs consumo energético medio por CCAA**

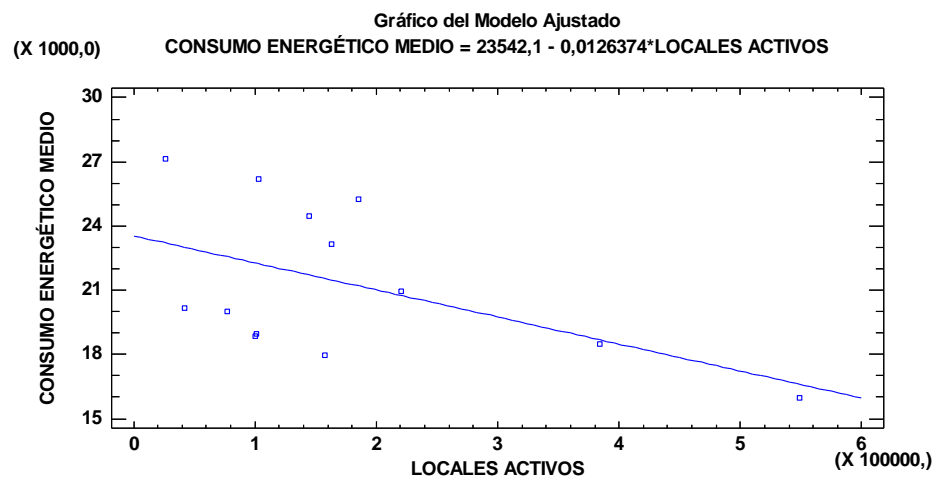


Ilustración 32, Locales Activos vs consumo energético medio por CCAA.

Al realizar el análisis bivalente se obtiene que el coeficiente de correlación es de -0,520232 por lo que la relación es indirecta y moderada.

- **PIB vs consumo energético medio por CCAA**

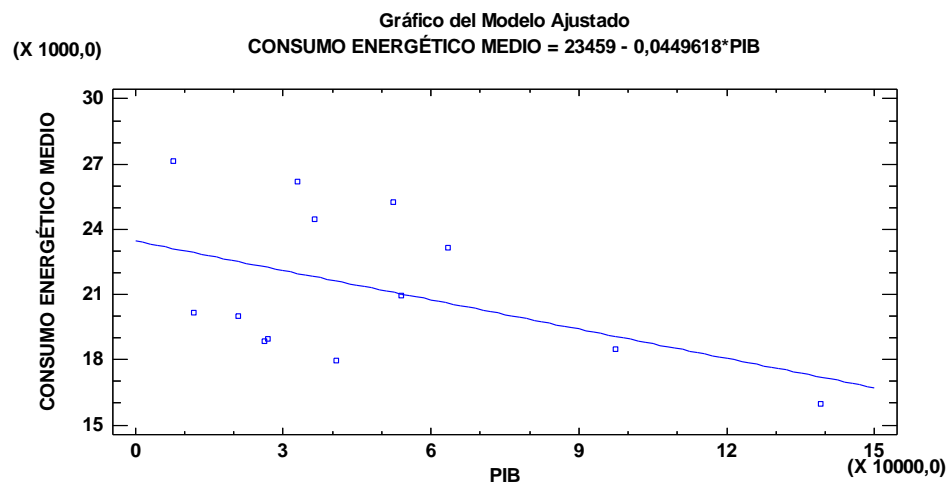


Ilustración 33, PIB vs consumo energético medio por CCAA.

Una vez realizado el análisis bivariante se obtiene que el coeficiente de correlación es de -0,46645, por lo que la relación es indirecta y baja.

- **Exportaciones vs consumo energético medio por CCAA**

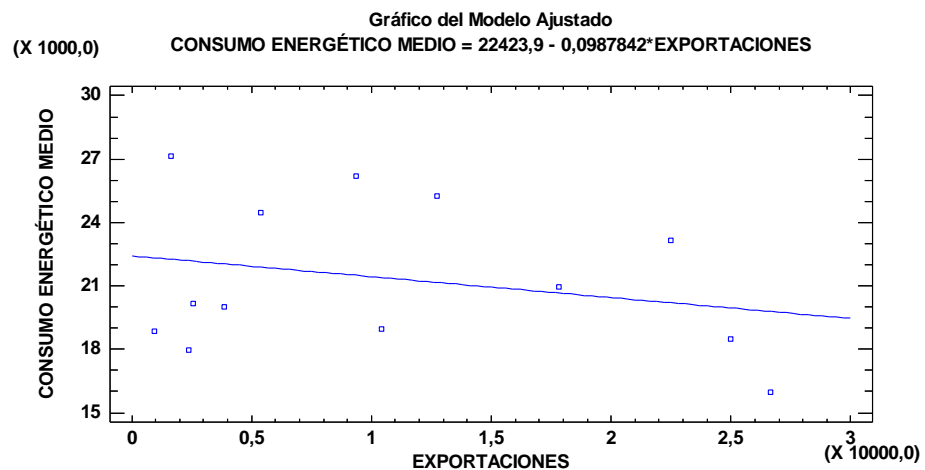


Ilustración 34, Exportaciones vs consumo energético medio por CCAA.

Tras realizar el análisis bivariante se obtiene que el coeficiente de correlación es de -0,261452 por lo que la relación es indirecta y muy baja.

- **Importaciones vs consumo energético medio por CCAA**

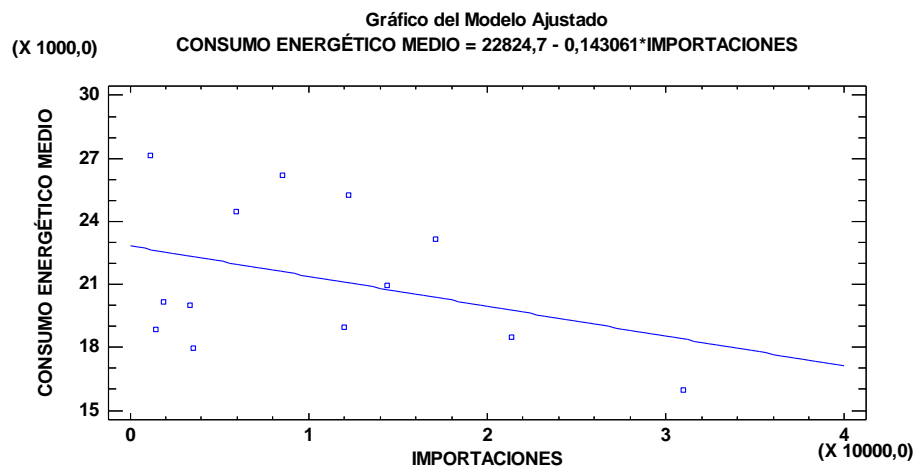


Ilustración 35, Importaciones vs consumo energético medio por CCAA

En este caso una vez realizado el análisis bivalente se obtiene que el coeficiente de correlación es de -0,36488 por lo que la relación es indirecta y baja.

- **Temperatura media vs consumo energético medio por CCAA**

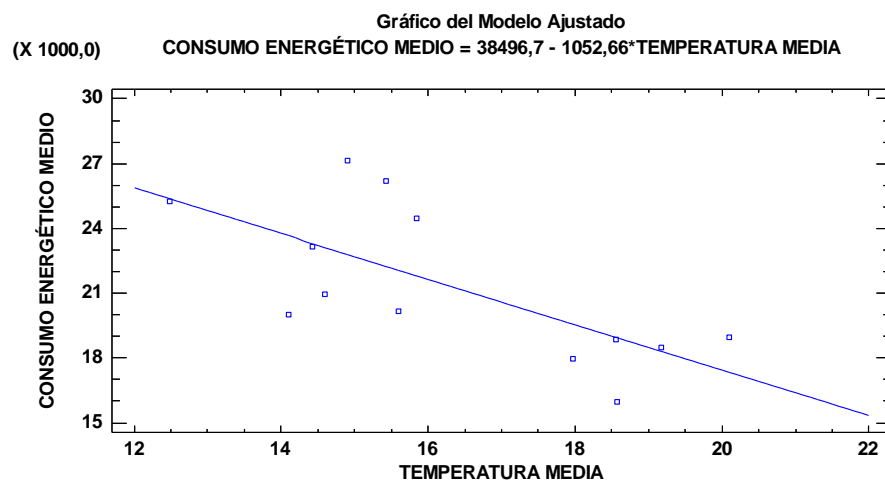


Ilustración 36, Temperatura Media vs consumo energético medio por CCAA.

Una vez realizado el análisis bivalente se obtiene que el coeficiente de correlación es de -0,695218 por lo que la relación es indirecta y moderada.

- **Horas de sol vs consumo energético medio por CCAA**

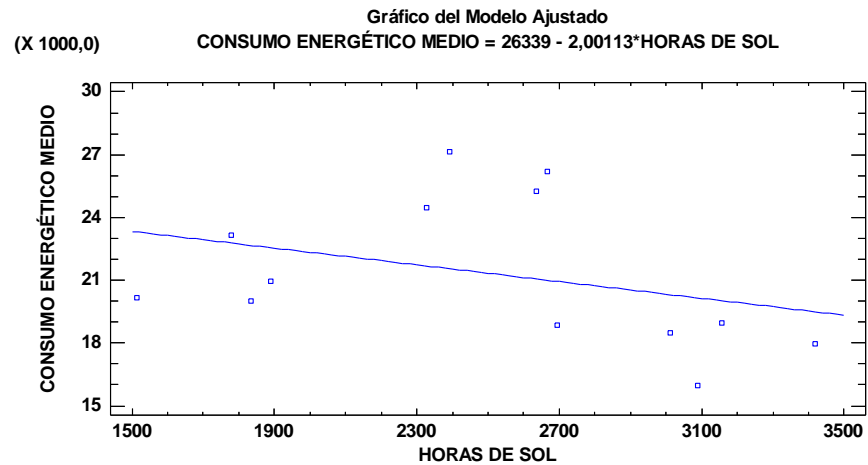


Ilustración 37, Horas de Sol vs consumo energético medio por CCAA.

Tras realizar el análisis bivariante se obtiene que el coeficiente de correlación es de -0,340605 por lo que la relación es indirecta y baja.

- **Emisiones CO₂ vs consumo energético medio por CCAA**

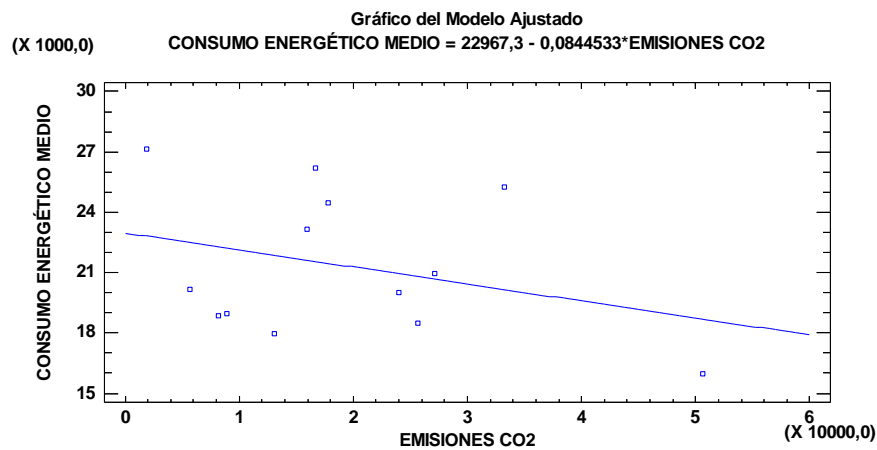


Ilustración 38, Emisiones CO₂ vs consumo energético medio por CCAA.

Una vez realizado el análisis bivariante se obtiene que el coeficiente de correlación es de -0,315442 por lo que la relación es indirecta y baja.

- Renta por hogar vs consumo energético medio por CCAA

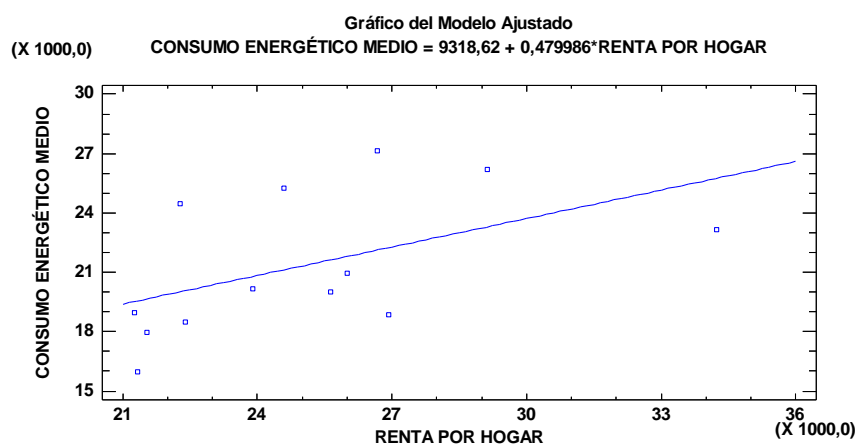


Ilustración 39, Renta por Hogar vs consumo energético medio por CCAA.

Una vez realizado el análisis bivalente se obtiene que el coeficiente de correlación es de 0,505569 por lo que la relaciones directa y moderada.

Como se puede apreciar una vez realizado el análisis bivalente al eliminar las comunidades citadas anteriormente, se observa que si que hay cambios, los cuales son positivos ya que aumentan las relaciones y el grado de estas.

4.3. MULTICOLINEALIDAD

Primera comprobación

- Matriz de correlación

	SUP	POB	EA	LA	PIB	EXP	IMP	TEMP	HS	CO2	RENT
SUP	1,00	0,56	0,52	0,52	0,51	0,36	0,47	-0,22	0,21	0,73	-0,27
POB	0,56	1,00	0,99	0,99	0,98	0,79	0,88	0,29	0,36	0,87	-0,31
EA	0,52	0,99	1,00	1,00	0,99	0,84	0,92	0,36	0,41	0,85	-0,32
LA	0,52	0,99	1,00	1,00	0,99	0,83	0,91	0,36	0,42	0,85	-0,34
PIB	0,51	0,98	0,99	0,99	1,00	0,88	0,94	0,31	0,37	0,84	-0,20
EXP	0,36	0,79	0,84	0,83	0,88	1,00	0,96	0,14	0,13	0,71	0,07
IMP	0,47	0,88	0,92	0,91	0,94	0,96	1,00	0,26	0,28	0,80	-0,11
TEMP	-0,22	0,29	0,36	0,36	0,31	0,14	0,26	1,00	0,68	-0,05	-0,52
HS	0,21	0,36	0,41	0,42	0,37	0,13	0,28	0,68	1,00	0,20	-0,54
CO2	0,73	0,87	0,85	0,85	0,84	0,71	0,80	-0,05	0,20	1,00	-0,24
RENT	-0,27	-0,31	-0,32	-0,34	-0,20	0,07	-0,11	-0,52	-0,54	-0,24	1,00

Tabla 13, Matriz de Correlación 1

Como se puede observar en la matriz de correlación hay muchos coeficientes por debajo de la diagonal principal superiores a 0,7, lo cual indica problemas de multicolinealidad.

- **Matriz inversa de correlación**

SUP	3,81	7,02	-31,44	24,97	-4,17	9,98	-8,83	4,23	-0,78	0,58	0,82
POB	7,02	739,38	912,29	-1670,52	-217,67	388,45	-260,81	93,78	62,27	99,74	-29,45
EA	-31,44	912,29	50339,65	-51965,61	1366,79	-994,23	-184,60	-351,17	580,96	180,12	-372,05
LA	24,97	-1670,52	-51965,61	54534,14	-1313,73	601,83	479,45	250,70	-647,23	-301,25	426,21
PIB	-4,17	-217,67	1366,79	-1313,73	274,63	-155,82	52,23	-37,40	-6,81	-18,68	-22,04
EXP	9,98	388,45	-994,23	601,83	-155,82	293,15	-192,39	73,26	18,13	61,82	-8,05
IMP	-8,83	-260,81	-184,60	479,45	52,23	-192,39	161,88	-47,86	-19,81	-49,87	12,73
TEMP	4,23	93,78	-351,17	250,70	-37,40	73,26	-47,86	24,56	0,96	19,06	0,39
HS	-0,78	62,27	580,96	-647,23	-6,81	18,13	-19,81	0,96	13,16	8,64	-5,34
CO2	0,58	99,74	180,12	-301,25	-18,68	61,82	-49,87	19,06	8,64	27,00	-4,47
RENT	0,82	-29,45	-372,05	426,21	-22,04	-8,05	12,73	0,39	-5,34	-4,47	8,52

Tabla 14, Matriz Inversa de Correlación 1.

Al observar la diagonal principal de la matriz inversa de correlación se observan coeficientes superiores a 10, lo cual también indica problemas de multicolinealidad.

- **Índice de acondicionamiento**

Valor Máx.	7,0061
Valor Mín.	0,000649467

$$I.C. = \sqrt{\frac{7,0061}{0,000649467}} = 103,8627$$

Tabla 15, Máximo y Mínimo 1.

Tras calcular el índice de acondicionamiento, se observa que es muy superior a 20, por lo que la multicolinealidad es elevada.

Tras realizar las tres pruebas y comprobar que existe el problema de multicolinealidad se va a proceder a eliminar una variable para así posteriormente volver a realizar las pruebas de multicolinealidad para comprobar si se ha resuelto el problema. El criterio a seguir para eliminar la variable va a ser el primero, ya que la diferencia entre los dos coeficientes más altos es mayor de 1.

Tras aplicar el primer criterio, se ha eliminado la variable **LOCALES ACTIVOS**, por lo que se va a proceder a volver a comprobar si existe problemas de multicolinealidad, pero sin esta variable.

Segunda comprobación

- Matriz de correlación

	SUP	POB	EA	PIB	EXP	IMP	TEMP	HS	CO ₂	RENT
SUP	1,00	0,56	0,52	0,51	0,36	0,47	-0,22	0,21	0,73	-0,27
POB	0,56	1,00	0,99	0,98	0,79	0,88	0,29	0,36	0,87	-0,31
EA	0,52	0,99	1,00	0,99	0,84	0,92	0,36	0,41	0,85	-0,32
PIB	0,51	0,98	0,99	1,00	0,88	0,94	0,31	0,37	0,84	-0,20
EXP	0,36	0,79	0,84	0,88	1,00	0,96	0,14	0,13	0,71	0,07
IMP	0,47	0,88	0,92	0,94	0,96	1,00	0,26	0,28	0,80	-0,11
TEMP	-0,22	0,29	0,36	0,31	0,14	0,26	1,00	0,68	-0,05	-0,52
HS	0,21	0,36	0,41	0,37	0,13	0,28	0,68	1,00	0,20	-0,54
CO ₂	0,73	0,87	0,85	0,84	0,71	0,80	-0,05	0,20	1,00	-0,24
RENT	-0,27	-0,31	-0,32	-0,20	0,07	-0,11	-0,52	-0,54	-0,24	1,00

Tabla 16, Matriz de Correlación 2.

Como se puede apreciar observando la matriz de correlación, hay muchos coeficientes por debajo de la diagonal principal de la matriz con valores superiores a 0,7 por lo que vuelve haber problemas de multicolinealidad.

- Matriz inversa de correlación

SUP	3,80	7,79	-7,64	-3,57	9,71	-9,05	4,11	-0,48	0,72	0,62
POB	7,79	688,20	-679,55	-257,91	406,89	-246,12	101,46	42,44	90,51	-16,39
EA	-7,64	-679,55	821,59	114,93	-420,75	272,27	-112,28	-35,78	-106,94	34,08
PIB	-3,57	-257,91	114,93	242,98	-141,32	63,78	-31,36	-22,40	-25,94	-11,77
EXP	9,71	406,89	-420,75	-141,32	286,51	-197,68	70,49	25,27	65,15	-12,75
IMP	-9,05	-246,12	272,27	63,78	-197,68	157,67	-50,06	-14,12	-47,22	8,98
TEMP	4,11	101,46	-112,28	-31,36	70,49	-50,06	23,41	3,94	20,44	-1,57
HS	-0,48	42,44	-35,78	-22,40	25,27	-14,12	3,94	5,48	5,07	-0,28
CO ₂	0,72	90,51	-106,94	-25,94	65,15	-47,22	20,44	5,07	25,34	-2,12
RENT	0,62	-16,39	34,08	-11,77	-12,75	8,98	-1,57	-0,28	-2,12	5,19

Tabla 17, Matriz Inversa de Correlación 2.

Al obtener la matriz inversa de correlación y observar la diagonal principal de la matriz, hay valores que son muy elevados a 10 lo cual indica problemas de multicolinealidad.

- Índice de acondicionamiento

Valor Máx.	6,02956
Valor Mín.	0,000532354

Tabla 18, Máximo y Mínimo 2.

$$I.C. = \sqrt{\frac{6,02956}{0,000532354}} = 106,4247$$

Tras calcular el índice de acondicionamiento, se concluye tras realizar las tres pruebas que la multicolinealidad es elevada.

El procedimiento para tratar de solucionar el problema de multicolinealidad va a ser el mismo que en la primera comprobación, es decir, se va a aplicar el primer criterio de eliminación de variable ya que la diferencia entre los dos coeficientes más altos es superior a 1.

En este caso la variable que se va a eliminar es **EMPRESAS ACTIVAS**.

Tercera comprobación

- Matriz de correlación

	SUP	POB	PIB	EXP	IMP	TEMP	HS	CO2	RENT
SUP	1,00	0,56	0,51	0,36	0,47	-0,22	0,21	0,73	-0,27
POB	0,56	1,00	0,98	0,79	0,88	0,29	0,36	0,87	-0,31
PIB	0,51	0,98	1,00	0,88	0,94	0,31	0,37	0,84	-0,20
EXP	0,36	0,79	0,88	1,00	0,96	0,14	0,13	0,71	0,07
IMP	0,47	0,88	0,94	0,96	1,00	0,26	0,28	0,80	-0,11
TEMP	-0,22	0,29	0,31	0,14	0,26	1,00	0,68	-0,05	-0,52
HS	0,21	0,36	0,37	0,13	0,28	0,68	1,00	0,20	-0,54
CO2	0,73	0,87	0,84	0,71	0,80	-0,05	0,20	1,00	-0,24
RENT	-0,27	-0,31	-0,20	0,07	-0,11	-0,52	-0,54	-0,24	1,00

Tabla 19, Matriz de Correlación 3

Tras obtener la matriz de correlación, sigue habiendo muchos coeficientes mayores a 0,7, los cuales generan problemas de multicolinealidad.

- Matriz inversa de correlación

SUP	3,73	1,47	-2,50	5,79	-6,52	3,07	-0,81	-0,28	0,94
POB	1,47	126,13	-162,85	58,88	-20,92	8,58	12,85	2,06	11,80
PIB	-2,50	-162,85	226,90	-82,47	25,70	-15,66	-17,40	-10,98	-16,54
EXP	5,79	58,88	-82,47	71,04	-58,24	12,99	6,95	10,38	4,70
IMP	-6,52	-20,92	25,70	-58,24	67,44	-12,85	-2,26	-11,78	-2,31
TEMP	3,07	8,58	-15,66	12,99	-12,85	8,07	-0,95	5,83	3,08
HS	-0,81	12,85	-17,40	6,95	-2,26	-0,95	3,92	0,41	1,20
CO2	-0,28	2,06	-10,98	10,38	-11,78	5,83	0,41	11,42	2,32
RENT	0,94	11,80	-16,54	4,70	-2,31	3,08	1,20	2,32	3,77

Tabla 20, Matriz Inversa de Correlación 3.

Al observar la matriz inversa de correlación se observan en la diagonal principal coeficientes superiores a 10, los cuales también indican problemas de multicolinealidad.

- **Índice de acondicionamiento**

Valor Máx.	5,06325
Valor Mín.	0,00254804

$$\sqrt{\frac{5,06325}{0,00254804}} = 44,5770$$

Tabla 21, Máximo y Mínimo 3.

Por último, al calcular el índice de acondicionamiento este es mayor que 20, por lo que la multicolinealidad es excesiva.

Como en las comprobaciones anteriores, se va a proceder a eliminar alguna variable para ver si así se soluciona el problema de la multicolinealidad. Al observar los coeficientes de la diagonal principal de la matriz inversa de correlación, la diferencia entre los dos coeficientes más altos es mayor que 1, por lo que se va a utilizar el criterio número 1 para eliminar la una de las variables.

En este caso la variable a eliminar es la de **PIB**.

Cuarta comprobación

- **Matriz de correlación**

	SUP	POB	EXP	IMP	TEMP	HS	CO2	RENT
SUP	1,00	0,56	0,36	0,47	-0,22	0,21	0,73	-0,27
POB	0,56	1,00	0,79	0,88	0,29	0,36	0,87	-0,31
EXP	0,36	0,79	1,00	0,96	0,14	0,13	0,71	0,07
IMP	0,47	0,88	0,96	1,00	0,26	0,28	0,80	-0,11
TEMP	-0,22	0,29	0,14	0,26	1,00	0,68	-0,05	-0,52
HS	0,21	0,36	0,13	0,28	0,68	1,00	0,20	-0,54
CO2	0,73	0,87	0,71	0,80	-0,05	0,20	1,00	-0,24
RENT	-0,27	-0,31	0,07	-0,11	-0,52	-0,54	-0,24	1,00

Tabla 22, Matriz de Correlación 4.

Una vez obtenida la matriz de correlación, se pueden observar coeficientes superiores a 0,7 por debajo de la diagonal principal de la matriz, esto hace indicar problemas de multicolinealidad.

- **Matriz inversa de correlación**

SUP	3,70	-0,33	4,88	-6,23	2,89	-1,01	-0,40	0,76
POB	-0,33	9,25	-0,30	-2,48	-2,65	0,36	-5,82	-0,07
EXP	4,88	-0,30	41,07	-48,91	7,30	0,63	6,39	-1,31
IMP	-6,23	-2,48	-48,91	64,53	-11,08	-0,29	-10,54	-0,44
TEMP	2,89	-2,65	7,30	-11,08	6,98	-2,15	5,07	1,94
HS	-1,01	0,36	0,63	-0,29	-2,15	2,59	-0,43	-0,07
CO2	-0,40	-5,82	6,39	-10,54	5,07	-0,43	10,89	1,52
RENT	0,76	-0,07	-1,31	-0,44	1,94	-0,07	1,52	2,57

Tabla 23, Matriz Inversa de Correlación 4.

Tras calcular la matriz inversa de correlación, existen varias variables con un coeficiente superior a 10 en la diagonal principal de la matriz, por lo que existe problema de multicolinealidad.

- **Índice de acondicionamiento**

Valor Máx.	4,12483
Valor Mín.	0,00932563

$$\sqrt{\frac{4,12483}{0,00932563}} = 21,0311$$

Tabla 24, Máximo y Mínimo 4.

Tras el cálculo del índice de acondicionamiento se puede concluir que la multicolinealidad es excesiva, ya que este está por encima de 20.

Para proceder a intentar arreglar el problema, se va hacer como en los casos anteriores, es decir, aplicando el primer criterio, ya que al observar la matriz inversa de correlación se observa que los dos coeficientes más altos la diferencia entre estos es mayor a 1, por lo que la variable a eliminar en este caso es **IMPORTACIONES**.

Quinta comprobación

- **Matriz de correlación**

	SUP	POB	EXP	TEMP	HS	CO2	RENT
SUP	1,00	0,56	0,36	-0,22	0,21	0,73	-0,27
POB	0,56	1,00	0,79	0,29	0,36	0,87	-0,31
EXP	0,36	0,79	1,00	0,14	0,13	0,71	0,07
TEMP	-0,22	0,29	0,14	1,00	0,68	-0,05	-0,52
HS	0,21	0,36	0,13	0,68	1,00	0,20	-0,54
CO2	0,73	0,87	0,71	-0,05	0,20	1,00	-0,24
RENT	-0,27	-0,31	0,07	-0,52	-0,54	-0,24	1,00

Tabla 25, Matriz de Correlación 5.

Como se puede apreciar en la matriz de correlación, el problema de multicolinealidad persiste, ya que hay variables relacionadas cuyo coeficiente es mayor a 0,7 lo cual indica problema de multicolinealidad.

- **Matriz inversa de correlación**

SUP	3,10	-0,57	0,16	1,82	-1,03	-1,41	0,71
POB	-0,57	9,16	-2,18	-3,08	0,35	-6,22	-0,09
EXP	0,16	-2,18	4,00	-1,10	0,41	-1,60	-1,64
TEMP	1,82	-3,08	-1,10	5,08	-2,20	3,26	1,87
HS	-1,03	0,35	0,41	-2,20	2,59	-0,48	-0,07
CO2	-1,41	-6,22	-1,60	3,26	-0,48	9,17	1,44
RENT	0,71	-0,09	-1,64	1,87	-0,07	1,44	2,56

Tabla 26, Matriz Inversa de Correlación 5.

Continuando con las pruebas, se puede apreciar en la diagonal principal que los coeficientes son inferiores a 10, por lo que esto significa que no hay problemas de multicolinealidad.

- **Índice de acondicionamiento**

Valor Máx.	3,31274
Valor Mín.	0,0575081

$$\sqrt{\frac{3,31274}{0,0575081}} = 7,5897$$

Tabla 27, Máximo y Mínimo 5.

Al calcular el índice de acondicionamiento, esta prueba indica que el modelo no tiene problemas de multicolinealidad, pero realmente si que tiene ya que en la matriz de correlaciones si que se ha detectado este problema.

En este caso el procedimiento para solucionar el problema va a ser basándose en el criterio 2 ya que la diferencia entre los dos coeficientes de la diagonal principal de la matriz inversa de correlación es menor a 1. Para ello se van a realizar dos regresiones diferentes, utilizando en cada una de ellas una variable diferente de las que generan el problema, tras esta regresión nos quedaremos con la variable que ofrezca un R² mayor. Tras realizar las regresiones se obtiene lo siguiente:

Regresión con la variable población, R²= 76,84.

Regresión con la variable emisiones de CO₂, R²= 90,38.

Por lo que la variable a eliminar va a ser **POBLACIÓN**.

Sexta comprobación

- **Matriz de correlación**

	SUP	EXP	TEMP	HS	CO2	RENT
SUP	1,00	0,36	-0,22	0,21	0,73	-0,27
EXP	0,36	1,00	0,14	0,13	0,71	0,07
TEMP	-0,22	0,14	1,00	0,68	-0,05	-0,52
HS	0,21	0,13	0,68	1,00	0,20	-0,54
CO2	0,73	0,71	-0,05	0,20	1,00	-0,24
RENT	-0,27	0,07	-0,52	-0,54	-0,24	1,00

Tabla 28, Matriz de Correlación 6.

Una vez obtenida la matriz de correlación se aprecia que por debajo de la diagonal principal de la matriz continúan habiendo coeficientes superiores a 0,7, por lo que aún sigue teniendo problemas de multicolinealidad el modelo.

- **Matriz inversa de correlación**

SUP	3,06	0,02	1,63	-1,01	-1,80	0,71
EXP	0,02	3,48	-1,83	0,49	-3,08	-1,66
TEMP	1,63	-1,83	4,05	-2,09	1,17	1,84
HS	-1,01	0,49	-2,09	2,58	-0,24	-0,06
CO2	-1,80	-3,08	1,17	-0,24	4,94	1,38
RENT	0,71	-1,66	1,84	-0,06	1,38	2,56

Tabla 29, Matriz Inversa de Correlación 6.

Observando la diagonal principal de la matriz inversa de correlación se puede comprobar que ningún coeficiente es superior a 10, esto indica que no hay problemas de multicolinealidad.

- **Índice de acondicionamiento**

Valor Máx.	2,49835
Valor Mín.	0,102657

$$\sqrt{\frac{2,49835}{0,102657}} = 4,9332$$

Tabla 30, Máximo y Mínimo 6.

El índice de acondicionamiento para este caso es menor de 10, por lo que no existe problema de multicolinealidad.

Tras realizar las tres pruebas se concluye que, si que hay problemas de multicolinealidad, para intentar resolverlo se va a aplicar el criterio de eliminación de variables número 2, ya que la diferencia entre los dos coeficientes más altos de la diagonal principal de la matriz inversa de correlación es menor a 1.

Tras resolver las regresiones múltiples empleando en cada una de ellas las distintas variables que generan el problema de multicolinealidad se obtiene que:

Regresión con la variable superficie, $R^2 = 65,31$.

Regresión con la variable emisiones de CO₂, $R^2 = 74,88$.

Por lo que se decide eliminar la variable **SUPERFICIE**.

Séptima comprobación

- Matriz de correlación

	CO2	EXP	TEMP	HS	RENT
CO2	1,00	0,71	0,14	0,13	0,07
EXP	0,71	1,00	-0,05	0,20	-0,24
TEMP	0,14	-0,05	1,00	0,68	-0,52
HS	0,13	0,20	0,68	1,00	-0,54
RENT	0,07	-0,24	-0,52	-0,54	1,00

Tabla 31, Matriz de Correlación 7.

Como se puede observar en la matriz de correlación en este caso solo existe correlación entre dos variables, por lo que aún existe multicolinealidad en el modelo.

- Matriz inversa de correlación

CO2	3,48	-3,07	-1,85	0,50	-1,67
EXP	-3,07	3,88	2,13	-0,83	1,80
TEMP	-1,85	2,13	3,18	-1,55	1,46
HS	0,50	-0,83	-1,55	2,24	0,17
RENT	-1,67	1,80	1,46	0,17	2,40

Tabla 32, Matriz Inversa de Correlación 7.

Tras observar la matriz inversa de correlación se puede apreciar que ningún coeficiente de la diagonal principal es superior a 10, por lo que no hay problemas de multicolinealidad.

- Índice de acondicionamiento

Valor Máx.	2,26998
Valor Mín.	0,102819

$$\sqrt{\frac{2,26998}{0,102819}} = 4,6986$$

Tabla 33, Máximo y Mínimo 7.

Una vez calculado el índice de acondicionamiento, este indica que no hay problemas de multicolinealidad, ya que este es menor a 10.

Para resolver el problema de multicolinealidad se va a llevar a cabo empleando el criterio de eliminación de variables número 2.

Tras resolver las regresiones se obtiene que:

Regresión con la variable exportaciones, $R^2 = 61,97$.

Regresión con la variable emisiones de CO₂, $R^2 = 71,81$.

Por lo que la variable a eliminar va a ser **EXPORTACIONES**.

Octava comprobación

- Matriz de correlación

	CO2	TEMP	HS	RENT
CO2	1,00	-0,05	0,20	-0,24
TEMP	-0,05	1,00	0,68	-0,52
HS	0,20	0,68	1,00	-0,54
RENT	-0,24	-0,52	-0,54	1,00

Tabla 34, Matriz de Correlación 8.

Tras calcular la matriz de correlación se observa que ningún coeficiente por debajo de la diagonal principal de la matriz es superior a 0,7 por lo que no hay problemas de multicolinealidad.

- Matriz inversa de correlación

CO2	1,18	0,50	-0,39	0,33
TEMP	0,50	2,20	-1,28	0,58
HS	-0,39	-1,28	2,17	0,41
RENT	0,33	0,58	0,41	1,60

Tabla 35, Matriz Inversa de Correlación 8

En la diagonal principal de esta matriz se puede observar que ningún coeficiente es superior a 10, por lo que no hay problemas de multicolinealidad.

- Índice de acondicionamiento

Valor Máx.	2,20463
Valor Mín.	0,273876

$$\sqrt{\frac{2,20463}{0,273876}} = 2,8372$$

Tabla 36, Máximo y Mínimo 8.

Una vez calculado el índice de acondicionamiento, se observa que este es menor a 10, por lo que no hay problemas de multicolinealidad.

Tras esta última comprobación se concluye que el modelo ya está libre de multicolinealidad, ya que ninguna de las tres pruebas tiene que dar positivo.

4.4. ANÁLISIS MULTIVARIANTE

Llegados a este punto, el último de los análisis a realizar, donde a partir de la proposición tanto del modelo teórico como el ajustado y explicación de cada parámetro, se va a proceder a validar el modelo, para ello el primer paso es comprobar la significatividad del modelo y variables, para ello se realizará mediante un contraste de hipótesis.

Seguidamente a través de gráficos y test se va a comprobar la normalidad de los residuos, en cuanto a gráficos se refiere se utilizará el histograma y gráfico probabilístico para comprobar la distribución de los datos y para el test se realizará un contraste de hipótesis con el test más restrictivo, el cual se explicará más hacia adelante.

A continuación, se realizará la comprobación de existencia o no existencia de heterocedasticidad en el modelo, esto se llevará a cabo a partir del contraste para la comprobación de heterocedasticidad.

Tras comprobar la heterocedasticidad se procederá a estudiar si el modelo tiene o no problemas de autocorrelación, a partir de los gráficos FAS y FAP y el test de Durbin Watson se podrá obtener conclusiones al respecto de este problema.

Por último, se obtendrán los puntos influyentes y anómalos del modelo, los cuales se obtendrán a partir de comprobar si cumplen con la condición para ser punto anómalo o influyente. En lo que a punto influyente se refiere puede ser a priori o posteriori, esto lo determinara la condición establecida para ello, las cuales están explicadas más hacia delante.

A continuación, el siguiente paso será resolver y arreglar los problemas que hayan podido surgir.

Para acabar se realizarán las predicciones. Para ello se va a realizar tres tipos de predicciones las cuales están explicadas posteriormente.

4.4.1 PROPOSICION MODELO TEÓRICO Y EXPLICACIÓN DE LOS PARÁMETROS

$$\text{CONSUMO ENERGÉTICO MEDIO} = \beta_0 + \beta_1 * \text{EMISIONES CO}_2 + \beta_2 * \text{TEMPERATURA MEDIA} + \\ + \beta_3 * \text{HORAS DE SOL} + \beta_4 * \text{RENTA POR HOGAR} + U$$

A continuación, se va a proceder a explicar cada parámetro.

- β_0

Valor medio de la variable CONSUMO ENERGÉTICO MEDIO cuando todas las variables explicativas valen 0.

- β_1

Incremento medio de la variable CONSUMO ENERGÉTICO MEDIO cuando se incrementa en una unidad la variable EMISIONES CO₂ y el resto de variables explicativas permanecen constantes.

- β_2

Incremento medio de la variable CONSUMO ENERGÉTICO MEDIO cuando se incrementa en una unidad la variable temperatura media y el resto de variables explicativas permanecen constantes.

- β_3

Incremento medio de la variable CONSUMO ENERGÉTICO MEDIO cuando se incrementa en una unidad la variable horas de sol y el resto de variables explicativas permanecen constantes.

- β_4

Incremento medio de la variable CONSUMO ENERGÉTICO MEDIO cuando se incrementa en una unidad la variable renta por hogar y el resto de variables explicativas permanecen constantes.

- U (PERTURBACIÓN)

Parte de los datos que el modelo no puede explicar con estas variables.

4.4.2. PROPOSICIÓN MODELO AJUSTADO

$$\text{CONSUMO ENERGÉTICO MEDIO} = 36151,3 - 0,112477 * \text{EMISIONES CO}_2 - \\ - 1451,63 * \text{TEMPERATURA MEDIA} + 2,83531 * \text{HORAS DE SOL} + 0,156752 * \text{RENTA POR HOGAR}$$

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	36151,3	10172,6	3,5538	0,0075
EMISIONES CO2	-0,112477	0,0546319	-2,05881	0,0735
TEMPERATURA MEDIA	-1451,63	421,363	-3,44507	0,0088
HORAS DE SOL	2,83531	1,62411	1,74576	0,1190
RENTA POR HOGAR	0,156752	0,225527	0,695049	0,5067

Análisis de Varianza

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	1,06887E8	4	2,67218E7	5,10	0,0244
Residuo	4,19529E7	8	5,24411E6		
Total (Corr.)	1,4884E8	12			

R-cuadrada = 71,8134 por ciento

R-cuadrado (ajustado para g.l.) = 57,7202 por ciento

Error estándar del est. = 2290,0

Error absoluto medio = 1444,55

Estadístico Durbin-Watson = 2,33334 (P=0,6509)

Autocorrelación de residuos en retraso 1 = -0,173805

Ilustración 40, Regresión Múltiple

4.4.3. SIGNIFICATIVIDAD

Llegados a este punto se va a comprobar la significatividad tanto de los parámetros como del modelo en general, para ello:

- Para el modelo:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_1: \text{Al menos uno es } \neq 0$$

- Para cada parámetro:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

Si $p. \text{ valor} > \alpha \rightarrow$ se acepta H_0 , por lo que no es significativo

Si $p. \text{ valor} < \alpha \rightarrow$ se rechaza H_0 , por lo que es significativo

A continuación, y como se ha mencionado anteriormente, se va a comprobar la significatividad tanto del modelo como de cada parámetro:

MODELO

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

P. Valor modelo: 0,0244

$$H_1: \text{Al menos uno es } \neq 0$$

α : 0,05

En este caso $P. \text{ Valor} < \alpha$, se rechaza H_0 por lo que el modelo es significativo.

CONSTANTE

$$H_0: \beta_0 = 0$$

P. Valor constante: 0,0075

$$H_1: \beta_0 \neq 0$$

α : 0,05

En este caso $P. \text{ Valor} < \alpha$, se rechaza H_0 por lo que la constante es significativa para el modelo.

EMISIONES CO₂

$$H_0: \beta_1 = 0$$

P. Valor Emisiones CO₂: 0,0735

$$H_1: \beta_1 \neq 0$$

α : 0,05

En este caso $P. \text{ Valor} > \alpha$, se acepta H_0 por lo que la variable emisiones CO₂ no es significativa para el modelo.

TEMPERATURA MEDIA

$$H_0: \beta_2 = 0$$

P. Valor Temperatura Media: 0,0088

$$H_1: \beta_2 \neq 0$$

α : 0,05

En este caso $P. \text{ Valor} < \alpha$, se rechaza H_0 por lo que la variable temperatura media es significativa para el modelo.

HORAS DE SOL

$$H_0: \beta_3 = 0$$

P. Valor Horas de Sol: 0,1190

$$H_1: \beta_3 \neq 0$$

α : 0,05

En este caso $P. \text{ Valor} > \alpha$, se acepta H_0 por lo que la variable horas de sol no es significativa para el modelo.

RENTA POR HOGAR

$$H_0: \beta_4 = 0$$

P. Valor Renta por Hogar: 0,5067

$$H_1: \beta_4 \neq 0$$

α : 0,05

En este caso $P. Valor > \alpha$, se acepta H_0 por lo que la variable renta por hogar no es significativa para el modelo.

Tras comprobar la significatividad tanto del modelo como de los distintos parámetros se observa que el modelo y la constante sí que son significativos, pero en el caso de las variables solo hay una que es significativa, temperatura media, por lo que posteriormente se procederá hacer cambios para intentar mejorar la significatividad.

4.4.4 NORMALIDAD DE RESIDUOS

En este punto se va a comprobar la normalidad de los residuos para ello se comprobará mediante los distintos gráficos y el contraste de hipótesis, para así poder sacar conclusiones respecto a la normalidad de los residuos.

- Histograma

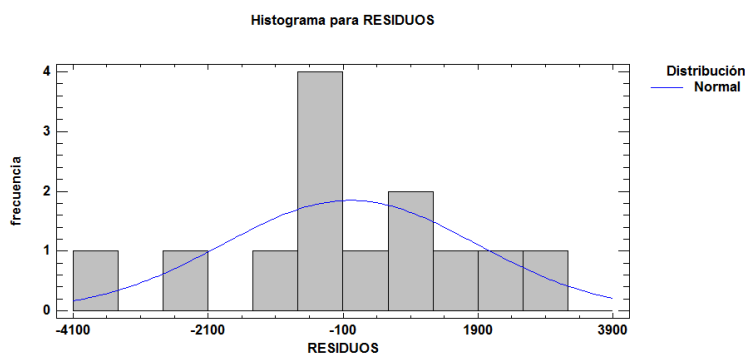


Ilustración 41, Histograma

Al observar el histograma, se puede observar que la distribución que siguen los residuos visualmente podría decantar que los residuos no son normales, ya que estos no siguen la forma de campana de Gauss la cual indica la normalidad de los residuos.

- **Grafico probabilístico**

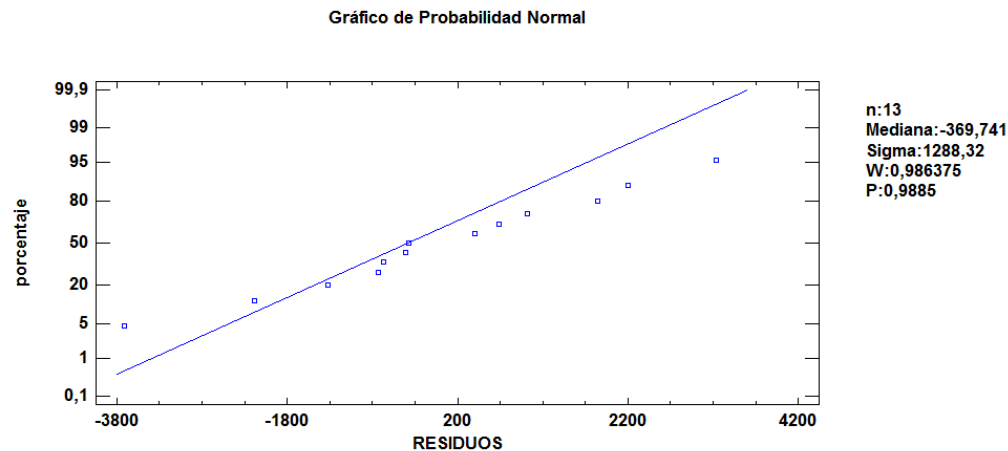


Ilustración 42, Gráfico Probabilístico

En lo que al gráfico probabilístico se refiere, se observa que la mayoría de los residuos están cerca de la línea de normalidad a excepción de cuatro puntos, por lo que se podría decir que los residuos en este caso si que parece ser que sigan una distribución normal.

- **Contraste de hipótesis**

Para la realización de este contraste hay que escoger el test más restrictivo, es decir el test más bajo y a partir de ahí:

H_0 : Los residuos son normales.

H_1 : Los residuos no son normales.

Pruebas de Normalidad para RESIDUOS

Prueba	Estadístico	Valor-P
Chi-Cuadrado	4,76923	0,781933
Estadístico W de Shapiro-Wilk	0,986375	0,988468
Valor-Z para asimetría	0,22421	0,82259
Valor-Z para curtosis	Datos Insuficientes	

Ilustración 23, Pruebas de Normalidad para Residuos.

Como se ha mencionado anteriormente al escoger el test más restrictivo (Chi-Cuadrado) y compararlo con alfa, se puede afirmar que los residuos son normales, ya que:

Si $P. \text{valor} \geq \alpha \rightarrow$ Se acepta H_0 , es decir, los residuos son normales.

En este caso el P. valor más restrictivo si que es mayor que alfa, por lo que se acepta la hipótesis nula.

4.4.5. HETEROCEDASTICIDAD

En este punto se va a comprobar la existencia o no existencia de heterocedasticidad en el modelo, pero para este análisis es necesario plantear un nuevo modelo previamente en el que los residuos se eleven al cuadrado tal que así:

- Modelo teórico

$$\text{Residuos}^2 = \beta_0 + \beta_1 * \text{EMISIONES CO}_2 + \beta_2 * \text{TEMPERATURA MEDIA} + \beta_3 * \text{HORAS DE SOL} + \beta_4 * \text{RENTA POR HOGAR} + U$$

- Modelo ajustado

$$\text{Residuos}^2 = 36151,3 - 0,112477 * \text{EMISIONES CO}_2 - 1451,63 * \text{TEMPERATURA MEDIA} + 2,83531 * \text{HORAS DE SOL} + 0,156752 * \text{RENTA POR HOGAR}$$

Tras este planteamiento del modelo, se va a realizar el análisis de heterocedasticidad para cada variable para ello se va a utilizar un contraste de hipótesis con el cual se comprobará si hay problemas de heterocedasticidad.

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
CONSTANTE	1,11721E7	1,13567E7	0,983742	0,3580
TEMPERATURA MEDIA	-393563,	485249,	-0,811053	0,4440
HORAS DE SOL	873,864	2052,76	0,425702	0,6831
EMISIONES CO2	-38,6408	64,4825	-0,599245	0,5679
RENTA POR HOGAR	-178,695	253,55	-0,704774	0,5037

Ilustración 44, Variables Explicativas.

$H_0: \sigma^2 = \text{constante}$ La variable no genera heterocedasticidad

$H_1: \sigma^2 \neq \text{constante}$ La variable genera heterocedasticidad

Si $P. \text{Valor} < \alpha \rightarrow$ Se rechaza H_0

- Temperatura Media

P. Valor: 0,4440 α : 0,05

En este caso como $P. \text{Valor} > \alpha$, se acepta H_0 , es decir, la variable no genera heterocedasticidad.

- Horas de Sol

P. Valor: 0,6831 α : 0,05

En este caso como P. Valor $> \alpha$, se acepta H_0 , es decir, la variable no genera heterocedasticidad.

- Emisiones CO₂

P. Valor: 0,5679 α : 0,05

En este caso como P. Valor $> \alpha$, se acepta H_0 , es decir, la variable no genera heterocedasticidad.

- Renta por Hogar

P. Valor: 0,5037 α : 0,05

En este caso como P. Valor $> \alpha$, se acepta H_0 , es decir, la variable no genera heterocedasticidad.

4.4.6 AUTOCORRELACIÓN

- **Funcion de autocorrelacion simple**

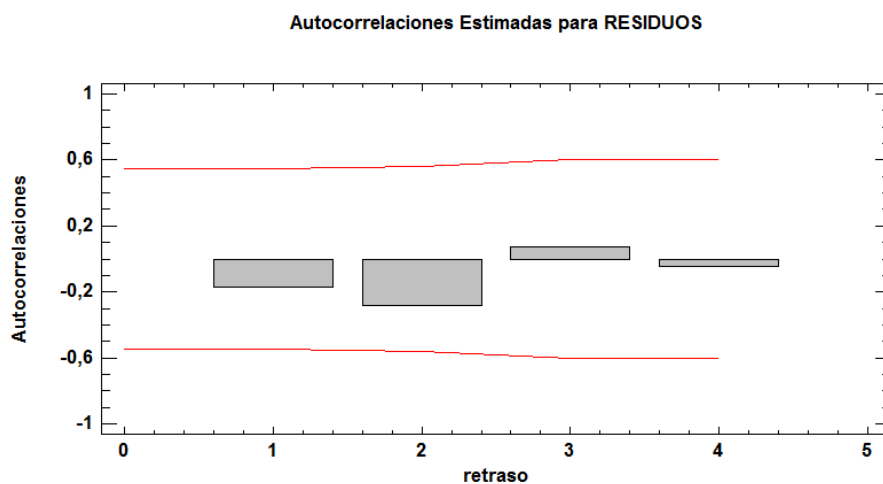


Ilustración 45, Función de Autocorrelación Simple.

- **Funcion de autocorrelacion parcial**

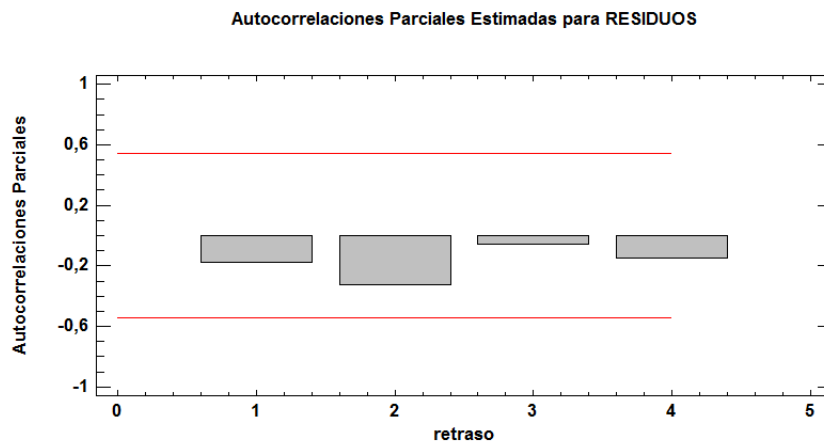


Ilustración 46, Función de Autocorrelación Parcial.

Tras observar los gráficos FAS y FAP, se puede apreciar que ningún factor de correlación alcanza los límites, lo cual podría indicar que el modelo no tiene problemas de autocorrelación. Pero para estar seguros de ellos, a continuación, se procederá a realizar el estadístico de Durbin Watson, con el que se tomará la conclusión final acerca del problema de autocorrelación.

- **Durbin Watson**

A partir del contraste de Durbin Watson se va a comprobar si el modelo tendría problema de autocorrelación de 1º orden, para ello:

$H_0: \rho_1 = 0$ no existe autocorrelación de 1º orden

$H_1: \rho_1 \neq 0$ existe autocorrelación de 1º orden

$\alpha: 0,05$

R-cuadrada = 71,8134 por ciento
 R-cuadrado (ajustado para g.l.) = 57,7202 por ciento
 Error estándar del est. = 2290,0
 Error absoluto medio = 1444,55
 Estadístico Durbin-Watson = 2,33334 (P=0,6509)

Ilustración 47, Estadístico Durbin-Watson.

Como se puede observar, el Estadístico Durbin-Watson es mayor que α , por lo que se acepta la hipótesis nula y se concluye que el modelo no tiene problemas de autocorrelación de 1º orden.

4.4.7. PUNTOS ANÓMALOS E INFLUYENTES

Llegados a este punto se van a analizar los puntos anómalos e influyentes del modelo. Para ello, en cada apartado se explica la condición para que sean de un tipo u otro. Como se ha mencionado anteriormente los puntos influyentes pueden ser a priori o posteriori. Dicho esto se va a proceder con el análisis:

- Puntos influyentes

Puntos Influyentes

		<i>Distancia de</i>	
<i>Fila</i>	<i>Influencia</i>	<i>Mahalanobis</i>	<i>DFITS</i>
10	0,383436	5,92413	-2,2241

Influencia media de un solo punto = 0,384615

Ilustración 48, Puntos Influyentes.

A priori \rightarrow influencia $> 2\bar{n}$

$$0,383436 < 0,7692$$

A posteriori $\rightarrow |DFITS| \geq 2\sqrt{n}$

$$2,2241 \geq 1,2403$$

Tras calcular los puntos influyentes y comprobar si es a priori o a posteriori, se concluye que es un punto influyente a posteriori una vez aplicados los criterios para ser un tipo u otro.

- Puntos anómalos

Residuos Atípicos

		<i>Y</i>		<i>Residuo</i>
<i>Fila</i>	<i>Y</i>	<i>Predicha</i>	<i>Residuo</i>	<i>Estudentizado</i>
10	17954,0	21663,2	-3709,25	-2,82

Ilustración 49, Residuos Atípicos.

R. Studentizado $> 2 \rightarrow$ punto anómalo.

R. Studentizado $> 3 \rightarrow$ Candidato a eliminar.

Al obtener los residuos atípicos del modelo, se observa que este al ser menor de 2, será punto anómalo. Este punto anómalo corresponde con la comunidad autónoma de Islas Canarias.

4.5. SOLUCIONES

Llegados a este punto se va a tratar de mejorar o solucionar los problemas que han surgido en el análisis del modelo para tratar de mejorarlo.

Tras realizar los distintos análisis se puede concluir que el único problema de este modelo es la significatividad de las variables del modelo, ya que ninguna es significativa, como se puede observar en la siguiente tabla:

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	36151,3	10172,6	3,5538	0,0075
EMISIONES CO2	-0,112477	0,0546319	-2,05881	0,0735
TEMPERATURA MEDIA	-1451,63	421,363	-3,44507	0,0088
HORAS DE SOL	2,83531	1,62411	1,74576	0,1190
RENTA POR HOGAR	0,156752	0,225527	0,695049	0,5067

Análisis de Varianza

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	1,06887E8	4	2,67218E7	5,10	0,0244
Residuo	4,19529E7	8	5,24411E6		
Total (Corr.)	1,4884E8	12			

R-cuadrada = 71,8134 por ciento

R-cuadrado (ajustado para g.l.) = 57,7202 por ciento

Error estándar del est. = 2290,0

Error absoluto medio = 1444,55

Estadístico Durbin-Watson = 2,33334 (P=0,6509)

Autocorrelación de residuos en retraso 1 = -0,173805

Ilustración 50, Regresión Múltiple.

Tras realizar el análisis de significatividad tanto del modelo como de las distintas variables se observa que tanto el modelo como la constante y la variable temperatura media resultan significativos, para intentar mejorar el modelo y tener así más datos significativos, con la ayuda del análisis de puntos anómalos e influyentes se va a proceder a eliminarlos y comprobar si el comportamiento del modelo sufre algún cambio.

Para ello se va a volver a realizar otra regresión múltiple, pero en este caso se eliminará del análisis Islas Canarias que es el punto anómalo que se obtuvo en el análisis.

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	39029,3	7510,04	5,19695	0,0013
RENTA POR HOGAR	0,0719925	0,167669	0,429372	0,6806
EMISIONES CO2	-0,154461	0,0426415	-3,62232	0,0085
HORAS DE SOL	4,68816	1,35747	3,45362	0,0106
TEMPERATURA MEDIA	-1703,69	320,889	-5,30928	0,0011

Análisis de Varianza

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	1,16706E8	4	2,91764E7	10,40	0,0045
Residuo	1,9638E7	7	2,80544E6		
Total (Corr.)	1,36344E8	11			

R-cuadrada = 85,5967 por ciento

R-cuadrado (ajustado para g.l.) = 77,3662 por ciento

Error estándar del est. = 1674,94

Error absoluto medio = 1021,64

Estadístico Durbin-Watson = 2,90389 (P=0,9276)

Autocorrelación de residuos en retraso 1 = -0,459754

Ilustración 51, Regresión Múltiple Sin Punto Anómalo.

Tras realizar esta regresión con la solución propuesta se va a proceder a comprobar nuevamente la significatividad del modelo y de cada uno de los parámetros, pero antes se plasmará el modelo obtenido tras esta solución:

$$\text{CONSUMO ENERGÉTICO MEDIO} = 39029,3 - 0,154461 \cdot \text{EMISIONES CO}_2 - \\ -1703,69 \cdot \text{TEMPERATURA MEDIA} + 4,68816 \cdot \text{HORAS DE SOL} + 0,0719925 \cdot \text{RENTA POR HOGAR}$$

A continuación y como se ha mencionado anteriormente se va a volver a comprobar la significatividad tanto del modelo como de las distintas variables, recordando que:

Si $P \text{ valor} \geq \alpha \rightarrow$ Se acepta H_0 , por lo que resultaría no significativa.

Si $P \text{ valor} \leq \alpha \rightarrow$ Se rechaza H_1 , por lo que resultaría significativa.

- Significatividad del modelo:

$H_0: B_1=B_2=B_3=B_4=0$ $\alpha=0,05$

$H_1: \text{Algún } B_i \neq 0$ p. valor modelo: 0,045

Se rechaza H_0 , por lo que el modelo resulta significativo.

- Significatividad de las variables:

Constante

$H_0: B_0 = 0$ $\alpha = 0,05$

$H_1: B_0 \neq 0$ p. valor constante: 0,0013

Se rechaza H_0 , por lo que la constante resulta significativa para el modelo.

- Emisiones de CO₂

$H_0: B_1 = 0$ $\alpha = 0,05$

$H_1: B_1 \neq 0$ p. valor emisiones CO₂: 0,0085

Se rechaza H_0 , por lo que la variable emisiones de CO₂ resulta significativa para el modelo.

- Temperatura Media

$H_0: B_2 = 0$ $\alpha = 0,05$

$H_1: B_2 \neq 0$ p. valor temperatura media: 0,0011

Se rechaza H_0 , por lo que la variable temperatura media resulta significativa para el modelo.

- Horas de Sol

$H_0: B_3 = 0$ $\alpha = 0,05$

$H_1: B_3 \neq 0$ p. valor horas de sol: 0,0106

Se acepta H_0 , por lo que la variable horas de sol resulta significativa para el modelo.

- Renta por Hogar

$H_0: B_4 = 0$ $\alpha = 0,05$

$H_1: B_4 \neq 0$ p. valor renta por hogar: 0,6808

Se acepta H_0 , por lo que la variable renta por hogar resulta no significativa para el modelo.

Tras este paso, se observa que en lo que a significatividad se refiere, sí que ha cambiado positivamente el resultado e incluso mejora el modelo obteniendo un 85,59%.

Pero aun así, sigue habiendo un problema con la significatividad de la ilustración 51, con la variable renta por hogar por lo que se va a proceder a eliminar esta variable del modelo.

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	41858,4	3415,13	12,2568	0,0000
TEMPERATURA MEDIA	-1750,56	285,959	-6,1217	0,0003
HORAS DE SOL	4,62853	1,27965	3,61702	0,0068
EMISIONES CO2	-0,159673	0,0387376	-4,12191	0,0033

Análisis de Varianza

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	1,16189E8	3	3,87295E7	15,37	0,0011
Residuo	2,01553E7	8	2,51941E6		
Total (Corr.)	1,36344E8	11			

R-cuadrada = 85,2173 por ciento

R-cuadrado (ajustado para g.l.) = 79,6738 por ciento

Error estándar del est. = 1587,26

Error absoluto medio = 1093,19

Estadístico Durbin-Watson = 2,92424 (P=0,9532)

Autocorrelación de residuos en retraso 1 = -0,485708

Ilustración 52, Regresión Múltiple Sin Variable Renta Por Hogar.

Al eliminar esta variable se observa que el R-cuadrada del modelo baja a 85,21, siendo así un buen modelo, en el que no tiene ningún tipo de problema, ya que el problema de la significatividad se ha solventado. Un modelo el cual resulta:

$$\text{CONSUMO ENERGÉTICO MEDIO} = 41858,4 - 0,159673 \cdot \text{EMISIONES CO}_2 - \\ -1750,56 \cdot \text{TEMPERATURA MEDIA} + 4,62853 \cdot \text{HORAS DE SOL}$$

5. PREDICCIONES

- Primera predicción

Para esta primera predicción se tiene que elegir una Comunidad Autónoma, al residir en la Comunidad Valenciana se decide escoger esta para ver el resultado. En esta primera predicción se va a comprobar lo que se ajusta el modelo a la realidad, es decir se va a comparar el dato real de consumo eléctrico en la Comunidad Valenciana con el que debería ser según el programa.

AJUSTADO	LC PARA LA MEDIA INFERIOR	LC PARA LA MEDIA SUPERIOR
18132,2	16293,8	19970,7

Tabla 37, Intervalo Para la Media de la Primera Predicción.

El dato real es de: **18493 KWH/m²**

El dato que da el programa es de: **18132,2 KWH/m²**

$$[16293,8 \leq \text{Consumo Energético Medio} \leq 19970,7]_{95\%}$$

Como se puede observar, tras esta primera predicción el modelo supone que el consumo energético medio para la Comunidad Valenciana debería ser 18132,2 KWH/m² por lo que el modelo se ajusta muy bien a los datos, ya que, el dato real para esta comunidad es de 18493 KWH/m².

- Segunda predicción

Para esta segunda predicción lo que se va a realizar son distintos escenarios en función de variar los datos reales por datos hipotéticos para ver qué sucede.

La primera modificación va a ser la de aumentar dos grados más la temperatura media.

DATO REAL	DATO HIPOTÉTICO
19,17 Grados Centígrados	21,17 Grados Centígrados
Consumo Eléctrico: 18493 KWH/m²	Consumo Eléctrico: 16151,0 KWH/m²

Tabla 38, Primera Modificación.

Tras esta hipótesis se observa que el dato real y el dato hipotético están muy cerca, por lo que se podría concluir que el modelo se ajusta bien, pero aun así para mayor seguridad sobre esta conclusión, se va a proceder a calcular el intervalo para la media para así comprobar si es correcta esta conclusión.

El calcular el intervalo para la media permitirá comprobar si el dato real está dentro del intervalo, en este caso se concluiría que es correcta la conclusión realizada a simple vista.

LC PARA LA MEDIA INFERIOR	LC PARA LA MEDIA SUPERIOR
13290,7	19011,3

Tabla 39, Intervalo Para la Media de la Primera Modificación.

$$[13290,7 \leq \text{Consumo Energético Medio} \leq 19011,3]_{95\%}$$

Como se puede observar, el dato real si que está dentro del intervalo para la media, por lo que se concluye que el modelo si que se ajusta bien.

La segunda modificación va a ser aumentar las horas de sol.

DATO REAL	DATO HIPOTÉTICO
3.010,67 Horas de Sol	5.300 Horas de Sol
Consumo Eléctrico: 18493 KWH/m²	Consumo Eléctrico: 20.548,0 KWH/m²

Tabla 40, Segunda Modificación.

Como se puede apreciar a simple vista, entre el dato real y el dato hipotético existe una pequeña diferencia, la cual puede hacer dudar en este caso del buen ajuste del modelo a la realidad.

Para ello y como se ha realizado en la hipótesis anterior, se va a calcular el intervalo para la media, para así comprobar que el dato real este dentro de este intervalo y así poder confirmar que el modelo se ajusta bien a la realidad.

LC PARA LA MEDIA INFERIOR	LC PARA LA MEDIA SUPERIOR
15906,4	25189,6

Tabla 41, Intervalo Para la Media de la Segunda Modificación.

$$[15906,4 \leq \text{Consumo Energético Medio} \leq 25189,6]_{95\%}$$

Tras calcular el intervalo para la media y observar el dato real de esta comunidad, se puede concluir que en este caso que el modelo se ajusta bien a la realidad, ya que el dato real está dentro del intervalo para la media.

Por último, la tercera modificación va a ser reducir las emisiones.

DATO REAL	DATO HIPOTÉTICO
25.696 Kilo Toneladas	23.300 Kilo Toneladas
Consumo Eléctrico: 18493 KWH/m²	Consumo Eléctrico: 18419,9 KWH/m²

Tabla 42, Tercera Modificación.

Como en los apartados anteriores, al comparar el dato real con el dato hipotético, se observa una pequeña diferencia entre estos, esta pequeña diferencia puede hacer pensar que en este caso el modelo también se ajusta bien, pero aun así se va a proceder a calcular el intervalo para la media para así poder confirmarlo con seguridad.

LC PARA LA MEDIA INFERIOR	LC PARA LA MEDIA SUPERIOR
16293,8	19970,7

Tabla 43, Intervalo Para la Media de la Tercera Modificación.

$$[16293,8 \leq \text{Consumo Energético Medio} \leq 19970,7]_{95\%}$$

Tras obtener el intervalo para la media, se observa que el dato real de consumo si que está dentro del intervalo, por lo que el modelo se ajusta bien a la realidad en este caso.

- Tercera predicción

Esta predicción va a consistir en lo mismo que en la primera, pero en este caso se va a comprobar todas las Comunidades Autónomas y posteriormente comentar la situación en la que se encuentran, para ello, a diferencia de la primera predicción se va a realizar de manera diferente, es decir, se va a proceder a calcular el intervalo de confianza para la media y compararlo con el dato real de cada Comunidad Autónoma para por último observar la situación de las diferentes comunidades.

Para una correcta situación de la comunidad mejor dicho para comprobar que el modelo se ajusta bien, el dato real debería estar dentro del parámetro para la media calculado, en caso contrario debería aumentar o disminuir el consumo energético medio por comunidad, a continuación, se estudiara caso por caso para poder sacar conclusiones y ver como se ajusta el modelo a la realidad.

En caso de tener una **V** significara que el dato real está dentro del parámetro para la media calculado, por lo que el modelo se ajustara bien en ese caso a la realidad.

En caso de tener una **X** significará que el dato real de esa comunidad autónoma esta fuera del intervalo para la media calculado, por lo que en esa comunidad se debería aumentar o disminuir el consumo energético medio, esto querrá decir que el modelo en este caso no se ajusta bien a la realidad.

C. AUTÓNOMA	DATO REAL	LC MEDIA INFERIOR	LC MEDIA SUPERIOR	SITUACIÓN
ANDALUCÍA	15993	12568.4	18543.9	V
ARAGÓN	26204	22881.8	26161.9	X
ASTURIAS	20007	20091.3	23553.1	X
CANTABRIA	20183	18202,1	23105,3	V
CASTILLA LA MANCHA	24458	20957,4	23112,9	X
CASTILLA LEÓN	25257	23975,6	29840.6	V
COMUNIDAD VALENCIANA	18493	16293,8	19970,7	V
GALICIA	20932	18935,0	22513,7	V
ISLAS BALEARES	18846	18766,3	22345,0	V
LA RIOJA	27121	24319,1	28768,7	V
MURCIA	18947	17464,5	22254,6	V
PAÍS VASCO	23159	20632,8	23941,5	V

Tabla 44, Situación Comunidades.

Tras realizar las predicciones de todas las comunidades autónomas, se va a proceder a comentar la situación de cada una como se ha mencionado anteriormente.

- Las siguientes comunidades están en situación correcta, ya que el dato real de consumo está dentro del intervalo para la media, estas son:
Andalucía, Cantabria, Castilla León, Comunidad Valenciana, Galicia, Islas Baleares, La Rioja, Murcia, País Vasco.
- Por otro lado, a continuación, se van a citar aquellas comunidades que tienen un consumo excesivo, es decir, consumen más energía eléctrica que de la que deberían, según el modelo ya que el dato de consumo real supera el intervalo para la media. Estas son:
Aragón, Castilla la Mancha.
- Por último, se van a indicar aquellas comunidades las cuales tienen un consumo insuficiente, ya que el dato de consumo real de consumo energético medio está por debajo del intervalo para la media. Estas son:
Asturias.

6. CONCLUSIONES

Tras realizar los distintos análisis con sus respectivos cálculos se puede concluir que:

A priori:

- Las variables planteadas desde un principio están muy correlacionadas entre ellas, de un total de 11 variables, han resultado aptas 4 resultando un modelo con un R^2 de 71,81%
- En cuanto a la significatividad tanto del modelo como de los distintos parámetros, no ha sido del todo satisfactorio ya que, solo salía significativo el modelo, la constante y la variable temperatura media.
- En cuanto a la normalidad de los residuos observando el histograma se podía pensar que los residuos no eran normales debido a la distribución gráfica. En cuanto al gráfico probabilístico prácticamente todos los datos a excepción de unos pocos estaban cerca de la línea de normalidad y por último al realizar el contraste de hipótesis escogiendo el test más restrictivo, se pudo concluir que los residuos eran normales.
- Tras comprobar que las variables explicativas estaban libres de heterocedasticidad, se procedió a comprobar la autocorrelación. A través del FAS y FAP se observaba que ningún factor de autocorrelación sobrepasaba los límites además tras realizar el contraste de Durbin Watson se comprobó que el modelo no tenía problemas de autocorrelación.
- A partir de calcular los puntos anómalos e influyentes se procede a eliminar el punto anómalo del modelo para intentar mejorarlo.

A posteriori:

- Tras eliminar el punto anómalo, Islas Canarias, el modelo mejora hasta llegar a un R^2 de 85,59% pero tras volver a comprobar la significatividad continúa habiendo una variable no significativa, renta por hogar.
- Se decide eliminar esta variable del modelo y el R^2 finalmente es de 85,21, un modelo libre de todo tipo de problemas.
- Al realizar las hipótesis se obtiene que:
Si aumentase la temperatura media dos grados, el consumo eléctrico en la comunidad valenciana bajaría.
Si aumentaran las horas de sol a 5,300 el consumo eléctrico aumentaría.
Si disminuyeran las emisiones de CO₂ el consumo eléctrico descendería, aunque muy poco.
- Tras realizar las predicciones se observa que el modelo se ajusta muy bien a la realidad a excepción de:
Aragón y Castilla la Mancha que tienen un uso de consumo excesivo según el modelo
Cantabria que tiene un uso insuficiente según el modelo.

7. LÍNEAS FUTURAS DE INVESTIGACIÓN

Tras la realización de este proyecto, a continuación, se van a plasmar posibles líneas futuras de investigación para así enfocar el estudio de una manera diferente a lo realizado en este proyecto. Estas son:

- Búsqueda de datos más actuales, es decir, actualizar la base de datos con datos más recientes, ya que los datos utilizados en este proyecto han sido del 2014.
- Incluir en el análisis a Ceuta y Melilla.
- Realizar el proyecto a través de otra estructura como la semilogarítmica, logarítmica, inversa o logarítmica inversa. Ya que este proyecto se ha realizado a través de la estructura lineal.
- Realizar un estudio sobre Cataluña y Madrid (las cuales se eliminaron nada más empezar el proyecto).
- Resolver el problema de la multicolinealidad a través del método componentes principales.
- Búsqueda de nuevas variables que puedan aportar más información al modelo.

8. BIBLIOGRAFÍA

PÁGINAS WEB

<http://www.comercio.gob.es/es-ES/comercio-exterior/estadisticas-informes/PDF/estadisticas-comercio-exterior/Enero%202015/Informe%20de%20Comercio%20Exterior%20-%202015-01.pdf>, 06/01/2017.

http://www.camaravalencia.com/es-ES/informacion/economica/estadisticas_economicas/Documents/CV_en_cifras_2014.pdf, 06/01/2017.

http://www.mapama.gob.es/es/cambio-climatico/publicaciones/publicaciones/informe_turismo_tcm7-197792.pdf, 06/01/2017.

<http://www.datosmacro.com/pib/espana?anio=2015>, 06/01/2017.

https://www.ecologistasenaccion.org/IMG/pdf/informe_calidad_aire_2014.pdf, 06/01/2017.

<http://www.ine.es/daco/daco42/bme/c19.pdf>, 06/01/2017

<http://www.minetad.gob.es/energia/balances/Balances/LibrosEnergia>, 09/01/2017.

http://www.mapama.gob.es/es/calidad-y-evaluacion-ambiental/temas/sistema-espanol-de-inventario-sei-/resumenemisionesgeiporccaaserie1990-2015_tcm7-454960.pdf, 09/01/2017.

http://www.appa.es/descargas/ESTUDIO_APPA_14_WEB.pdf, 09/01/2017.

http://www.mapama.gob.es/es/alimentacion/temas/la-agricultura-ecologica/estadisticas_ae_2014_definitivopdf_tcm7-405122.pdf, 09/01/2017.

http://www.mapama.gob.es/es/calidad-y-evaluacion-ambiental/publicaciones/perfil_ambiental_2014.aspx, 09/01/2017.

<http://www.mapama.gob.es/es/calidad-y-evaluacion-ambiental/temas/sistema-espanol-de-inventario-sei-/>, 09/01/2017.

<http://www.ine.es/prensa/np924.pdf>, 12/01/2017.

<https://www.certicalia.com/blog/consumo-global-por-comunidades-autonomas>, 12/01/2017.

<http://www.ine.es/jaxiT3/Datos.htm?t=9949>, 12/01/2017.

LIBROS

Estadística Modelos y Métodos. 1, Fundamentos - Daniel Peña Sánchez de Rivera.

ECONOMETRÍA Segunda Edición – Alfonso Novales.

ECONOMETRÍA Tercera Edición – DAMODAR N. GUJARATI.

Estadística para administración y economía- Paul Newbold / William L. Carlson / Betty Thorne

APUNTS D'ESTADISTICA – Fortunato Crespo Abril.

APUNTES

Apuntes de la asignatura Econometría. Profesores Jorge Jordán y Bárbara Mico. Curso 14/15.